**초등 수학 학습 결손 추적 AI 모델 개발 보고서**

빅데이터 8기 해~조~팀

김솔미, 김영규, 박지석, 이현희, 허현강

**요약**

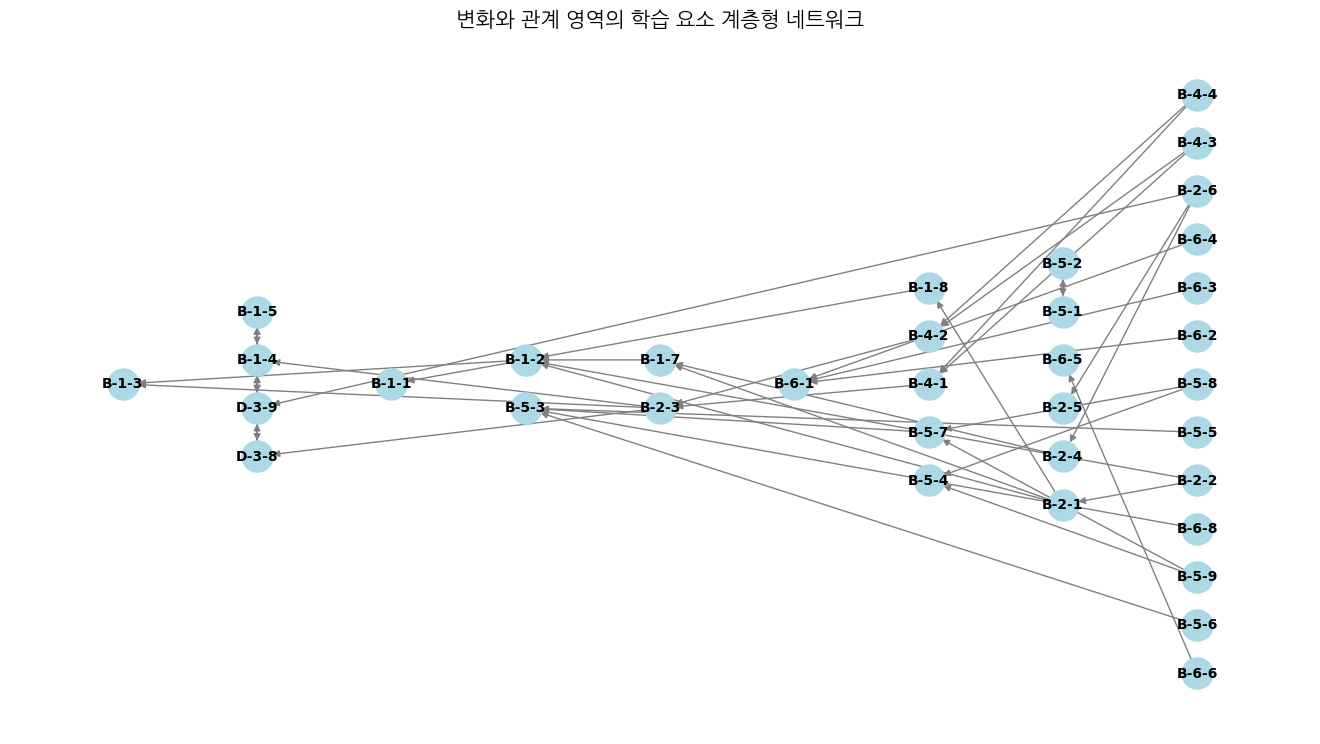
본 프로젝트는 Knowledge Space Theory (Stahl et al., 2022)[[1]](#footnote-1)와 Item Response Theory(IRT)를 바탕으로 초등학교 학생들의 수학 학습 결손을 추적하고, 학생 맞춤형 추천 학습을 제공하는 지능형 튜터링 시스템을 개발하는 것을 목표로 한다. 지능형 튜터링(박성익 & 임철일, 2022)[[2]](#footnote-2)이란 학습자의 문제해결 과정을 점검하고 오답을 분석하여 학습자에게 필요한 지원과 피드백을 제공하는 것을 말한다. 최근 교육 시장에서 기초학력(학교 교육과정을 통하여 갖추어야 하는 최소한의 성취기준을 충족하는 학력)[[3]](#footnote-3)향상과 수준별 맞춤형 학습의 필요성이 강조되고 있는 상황에서, 인공지능을 활용한 학습 진단과 자동화된 학습 프로그램은 교육 분야에서 중요한 역할을 담당할 수 있다. 8개의 feature와 1610개의 sample로 이루어진'label\_math\_ele' 데이터와, 6개의 feature와 642개의 sample로 이루어진 'education\_2022' 데이터, 7개의 feature와 1677358개의 sample로 이루어진 'merged\_df'의 세 데이터를 바탕으로 2022 개정 초등학교 수학과 교육과정을 계열화하여 선수학습 및 후속학습 추천 AI 모델을 개발하였다. 선수학습 추천 알고리즘은 이진 분류 문제를 해결하기 위해 설계된 EnhancedBinaryNN을 사용하여 학습 결손을 추척해, 맞춤형 선수학습 개념을 추천하였다. 후속학습 추천 알고리즘은 'Manhalanobis 거리 이론'을 기반으로 가장 가까운 거리에 있는 후속학습 개념을 추천하였다. 이는 교육부의 AI 디지털 교과서 도입 계획과 연계되어, 개별화된 맞춤형 학습을 실현함으로써 교육 격차를 해소하고 학습자가 자신의 강점을 강화할 수 있는 기회를 제공하는데 큰 역할을 할 것이다.

**Ⅰ. 서론**

**1) 이론적 배경:**

**A. KST**  
지식 공간 이론(Knowledge Space Theory, Doignon and Falmagne, 1999)[[4]](#footnote-4)은 특정 도메인에서 지식 구조를 수학적으로 운영하기 위해 집합 및 순서 이론적 틀을 제안하는 이론이다. 이 이론은 항목 간에 종속 관계가 존재한다고 가정하며, 특정 항목이나 항목의 하위 집합에 대한 지식이 더 어렵거나 복잡한 다른 항목에 대한 지식의 선행 조건이 될 수 있다고 본다. 이러한 모든 지식 상태의 집합을 '지식 구조(knowledge structure)'라고 한다.

이 이론을 바탕으로 2022 개정 초등학교 수학과 교육과정의 영역을 기준으로 복잡하게 구성된 개념들을 영역별, 내용 요소별로 [계열화](#라벨링_계열화방법)하여, 각 영역에서 학습자의 학습 상태를 추적하고 결손을 분석한다. 다음은 지식 공간 이론을 기반으로 세분화된 5가지 영역 중 변화와 관계에 대한 개념의 계열화를 나타낸 예시이다.



[그림 1. 변화와 관계 영여의 학습 요소 계층형 네트워크]

위 그래프는 '변화와 관계' 영역의 노드들을 계열화하여 연결한 것으로, 계층 구조를 따르며 학습의 선후 관계에 따라 노드들을 배치하였다. 왼쪽에서 오른쪽으로 진행되며, 상위 레벨의 노드는 오른쪽에 배치하였다. 상위 레벨의 학습 요소에서 시작하여 화살표를 따라가면서 하위 레벨로 이동할 수 있다. 이렇게 함으로써, 어떤 학습 요소가 먼저 이루어져야 하고, 다음 학습이 무엇인지를 시각적으로 쉽게 이해할 수 있다.

**B. IRT**   
IRT(Item Response Theory)는 평가 도구의 항목별 응답과 잠재 심리적 특성 간의 관계를 모델링하는 이론이다. 이 이론은 평가 항목의 성능을 상세히 분석하고, 이를 통해 평가 도구의 신뢰성과 타당성을 강화하는 데 유용하게 사용된다. IRT는 기존의 고전적 검사 이론(CTT)과 달리 각 항목의 응답 확률을 잠재 특성(θ)와 관련하여 수학적 모델로 설명한다. 사용한 데이터에서는 3모수 로지스틱 모델을 이용하여 문제의 난이도, 변별도, 추측도를 도출해 적재해놓았다.

● IRT 모델의 종류:

○ 1-모수 로지스틱 모델(1PL): 항목의 난이도만 고려하는 가장 단순한 모델로, 모든 항목의 변별력이 동일하다고 가정한다.

○ 2-모수 로지스틱 모델(2PL): 항목의 난이도와 변별력을 모두 고려하며, 각 항목이 개인의 능력을 얼마나 잘 구분하는지 평가할 수 있다.

○ 3-모수 로지스틱 모델(3PL): 난이도, 변별력, 추측 모수를 포함하여 피험자가 항목을 맞출 확률이 단순 추측에 의한 영향을 반영한다.

**C. Mahalanobis 거리**

Mahalanobis 거리는 다차원 공간에서 데이터 포인트 간 거리를 의미하며 특정 값 X가 데이터의 분포에서 얼마나 멀리 떨어져 있는지를 나태는 거리이다. 이 거리는 변수간 상관관계까지 고려했기 때문에, [[5]](#footnote-5)변수들이 상관관계를 갖거나 다른 크기를 가질 때더욱 정확한 결과를 제공한다. 공분산 행렬은 다차원 데이터에서 각 변수 간의 상관관계를 나타내는 행렬이다. 데이터의 여러 변수들이 어떻게 변하는지에 대한 정보를 제공한다.

**D. 학습 결손 보완의 중요성**

학습 결손은 학생들이 학습 과정에서 필수적인 지식을 습득하지 못하여 이후 학습에 어려움을 겪는 상황을 의미한다. 이러한 결손을 효과적으로 보완하는 것은 교육의 질을 높이고, 학습자 개개인의 잠재력을 최대한 발휘하도록 하는 데 중요한 역할을 한다. 특히, 학습 결손을 조기에 발견하고 체계적으로 보완하는 것은 학생들의 학습 성과 향상에 필수적이다. (정순환, 2015)[[6]](#footnote-6)은 학습 결손이 발생한 부분을 파악하여 개별화된 학습 프로그램을 계획하는 것이 필요하다고 강조하였다.

**E. 맞춤형 학습의 중요성과 최근 동향**

맞춤형 학습은 각 학생의 특성과 학습 속도에 맞춰 개인화된 학습 경험을 제공하는 것을 목표로 한다. 이는 학생의 자발적 학습을 촉진하고 학습 동기와 참여도를 높이는 데 효과적이다. (성지현, 2023)[[7]](#footnote-7)은 맞춤형 학습의 도입이 학습자의 강점을 강화하고 약점을 보완하여 전체적인 학습의 질을 향상시키는 데 기여할 것으로 기대하였다.

최근 기술의 발전은 교육 분야에서 맞춤형 학습의 실현 가능성을 높이고 있다. 학습진단 AI 기술은 학생들의 학습 데이터를 분석하여 개인별 학습 성취도를 평가하고, 학습 결손을 파악하여 맞춤형 학습 경로를 제공하는 데 중점을 두고 있다. 이러한 기술은 학생들이 자신의 학습 상태를 실시간으로 점검하고, 학습 효과를 극대화할 수 있도록 지원하는 중요한 역할을 한다

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **국내 기술 현황** | |  |  |  | | --- | --- | --- | | **제공** | **서비스 명** | **튜터모델(교수학습 전략)** | | 교육부 | 똑똑수학  탐험대 | 초기에는 교육 전문가들이 구성한 알고리즘에 따라 학습자 진단과 추천 서비스를 제공하며, 학습 이력이 충분히 축적되면 기계학습 모델을 활용하여 학습자의 지식 상태를 진단한다. 이를 통해 취약점을 강화할 수 있는 강화학습 기반의 학습 경로를 추천한다. | | 웅진 | 싱크빅 AI | 매 1/1000초마다 학습자별 상태를 기록하여 빅데이터를 누적, 현재 500억 건 이상의 데이터를 수집·분석하여 학생들의 학습 행동 패턴을 분석한다. 이 데이터를 기반으로 문항별 성과를 분석하여 학습 성취도를 예측하고, 실시간 학습 예측을 통해 필요한 문제만을 출제하는 맞춤형 커리큘럼을 제공한다. 또한 AI 분석을 통해 오답의 원인을 찾아 취약 개념을 집중 관리하며, 학부모에게는 AI 월간 분석지를 통해 학습 결과와 학습 습관을 제공한다. | | 천재  교육 | 닥터매쓰 | 140종 이상의 교재 메타데이터를 연동하여 유사 문항 및 쌍둥이 문항 추천, 채점 기능을 제공한다. 학생 개인의 취약점을 분석하고 추천 AI 엔진을 기반으로 수준별 맞춤 문제를 제공하는 수학 학습 서비스이다. 이를 위해 문항반응이론, BKT, DKT 모형 등 다양한 이론을 접목하여 AI를 활용하고 있다.[[8]](#footnote-8) 특히, 천재교육은 학습 취약 개념을 자동으로 진단하고 해당 개념을 학습할 수 있는 서비스를 제공하는 시스템에 대한 특허를 국내에서 출원하여 권리를 확보하였다(천재교육, 2023)[[9]](#footnote-9). | |
| **국외 기술 현황** | |  |  |  | | --- | --- | --- | | **제공** | **서비스 명** | **튜터모델(교수학습 전략)** | | YiXue | Squirrel AI | 먼저 학습 과정을 진단하여 학생의 지식 상태와 능력을 평가한다. 그 결과에 따라 학생에게 적합한 학습 자료와 개인화된 학습 경로를 제공하여, 이미 알고 있는 것에 시간을 낭비하지 않고 학습 효율을 향상시킨다. 또한 분석된 학습자의 특성을 바탕으로 지식 그래프에서 학습자의 출발점을 조정하여 맞춤형 학습을 구현한다. | | McGraw Hill | ALEKS | Knowledge Space Theory(KST)를 기반으로 학생들의 지식 상태를 평가하고 그에 맞는 학습 경로를 제시하는 시스템이다. 이 시스템은 학생들의 현재 학습 상태를 정확하게 파악하여, 학습할 준비가 된 주제를 결정하고 대화형 학습 모듈에서 선택할 수 있도록 지원한다. 이를 통해 학생들은 자신에게 적합한 학습을 효율적으로 진행할 수 있다.[[10]](#footnote-10) | |

[표 1. 국내외 학습진단 AI 기술 현황]

**F. 디지털 기반 교육 혁신과 AI의 역할**

교육 분야에서는 교과서와 학습 콘텐츠의 디지털화가 지속적으로 진행되어 왔으며, 평가 방식 또한 디지털 기술과 인공지능을 활용하는 방향으로 혁신되고 있다. 2007년 '디지털교과서 상용화 정책'의 시작으로 다양한 교과에서 디지털교과서의 개발이 이루어졌고(Jeong & Sung, 2017)[[11]](#footnote-11) , 최근에는 AI 디지털교과서의 개발과 도입이 계획되고 있다(교육부, 2023)[[12]](#footnote-12).

이러한 변화는 학습 데이터의 수집과 분석을 용이하게 만들어, 교사가 학생 개개인의 학습 상태를 정확히 파악하고 맞춤형 교육을 제공할 수 있는 기반을 마련한다. 따라서 학습 결손을 확인하고 적절한 선수/후속 학습을 추천하는 프로그램을 제공한다면, 공교육의 신뢰도를 높이고 학생들의 학습 효과를 극대화할 수 있을 것이다. 이는 교육의 질을 향상시키는 동시에, 우리 회사가 이러한 솔루션을 제공함으로써 수익을 창출할 수 있는 기회를 의미한다.

또한, 이러한 프로그램을 사교육 시장에 적용한다면, 개인화된 학습 솔루션을 찾는 학습자와 학부모의 요구를 충족시켜 기업의 매출 증대에도 긍정적인 영향을 미칠 것이다.

**G. 수학 교육의 결손 추적**

수학은 선수 학습이 필수적인 명확한 위계 구조를 가진 학문으로, 이전 단계의 학습이 제대로 이루어지지 않으면 다음 단계로의 진행이 어렵다. 이러한 특성 때문에 수학은 개별 학습 지도와 지식 전달 방식, 학습자 분석이 용이하여 인공지능을 통한 맞춤형 학습에 가장 적합한 교과로 평가된다(정제영, 2021)[[13]](#footnote-13). 따라서 수학 교육에서 학습 결손을 정확히 추적하고, 이에 따른 선수/후속 학습을 추천하는 프로그램의 개발은 매우 중요하다. 이러한 프로그램을 통해 학생들은 자신의 학습 상태를 정확히 파악하고 필요한 부분을 보완할 수 있으며, 교사는 효율적인 교육 전략을 수립할 수 있을 것이다.

**2) 프로그램 개발의 필요성**

최근 교육 시장에서는 기초학력 향상[[14]](#footnote-14)과 수준별 맞춤형 학습의 필요성이 더욱 강조되고 있다. 그러나 현재 학습 부진 학생을 진단하기 위한 평가 도구는 주로 이전 학년의 선수 학습 내용을 종합적으로 다루어, 학생들이 수학의 어떤 내용에서 구체적으로 결손이 발생했는지 파악하기 어렵다는 한계가 있다. 대부분의 학습 부진 학생들은 특정 영역에서의 선수 학습 결손으로 인해 학습에 어려움을 겪으며, 개인에 따라 결손 내용과 정도가 다양하다.

따라서 각 학생의 학습 부진 원인을 정확히 진단하고, 이에 따른 개별 지도를 제공하는 것이 바람직한 해결책으로 제시되고 있다(박성익, 1986)[[15]](#footnote-15). 이를 위해서는 학생들의 수학 학습 결손 영역을 정확히 진단할 수 있는 적절한 도구의 개발이 필수적이다.

또한, 학습 부진 내용을 효과적으로 보충하기 위해서는 각 단원별 학습 과제를 위계적으로 배열하고, 학습자들이 각 내용을 어느 정도 습득했는지 확인하며, 어떤 학습 과정에서 부진이 발생했는지를 구체적으로 밝히는 작업이 선행되어야 한다(김순택, 1979)[[16]](#footnote-16).

교육부는 AI 디지털 교과서와 같은 기술을 활용하여 개별 학생의 학습 결손을 실시간으로 분석하고, 이를 보완할 수 있는 맞춤형 학습 경로를 제시하는 방안을 도입하고 있다. 이러한 맞춤형 접근은 학생들이 자신에게 필요한 학습을 주도적으로 수행하도록 돕고, 교육 격차를 줄이는 데 기여할 것으로 기대된다.

이에 따라 학습 결손을 추적하는 AI 서비스의 개발은 기존 평가 도구의 한계를 극복하고, 학생에게 맞춤형 교육을 제공함으로써 학습 효율성과 성취도를 향상시키는 데 필수적이다. 우리 회사가 이러한 프로그램을 개발하여 공교육 현장에 제공한다면, 교육의 신뢰도를 높이는 동시에 기업의 수익 창출에도 도움이 될 것이다. 또한, 사교육 시장에서도 개인화된 학습 솔루션에 대한 수요가 높아지고 있어, 해당 프로그램을 적용함으로써 매출 증대에 긍정적인 영향을 미칠 수 있다.

따라서 학습 결손 추적 및 선수·후속 학습을 추천하는 AI 프로그램의 개발은 교육 분야의 혁신을 주도하고, 학생들의 학습 경험을 향상시키는 핵심적인 역할을 할 것으로 기대된다.

**3) 학습 결손 추적 AI 서비스의 순서도**

|  |
| --- |
|  |

[그림 2. 학습 결손 추척 AI 서비스의 순서도]

**Ⅱ. 본론**

**1) 데이터 소개 및 분석**

**A. label\_math\_ele 데이터 소개**

본 데이터는 AI Hub의 수학 분야 학습자 역량 측정 데이터 중 수학지식체계 데이터셋 라벨 JSON 파일 전처리한 데이터이다. ToConcept(선수학습)과 FromConcept(후속학습) 개념을 연결하여 이들 중 선수학습-후속학습이 '초등-초등' 혹은 '초등-중등' 인 경우만을 남겨두고 전처리하였다. 이 데이터는 8개의 feature와 1610개의 sample로 이루어져있다.

|  |  |  |
| --- | --- | --- |
| **feature** | **해석** | **예시** |
| from\_id | 후속단위 개념 id | 2615 |
| from\_semester | 후속단위개념의 ‘학교-학년-학기’ | 초등-초5-1학기 |
| from\_chapter\_name | 후속단위개념이 속하는 단원의 이름 | 자연수의 혼합 계산 > 덧셈과 뺄셈이 섞여 있는 식을 계산해 볼까요 |
| to\_id | 선수단위 개념 id | 7933 |
| to\_chapter\_name | 선수단위개념의 ‘학교-학년-학기’ | 덧셈과 뺄셈 > 세 수의 계산을 해 볼까요 |
| to\_semester | 선수단위개념이 속하는 단원의 이름 | 초등-초2-1학기 |
| from\_g | 후속단위 개념 학교 | 초등 |
| to\_g | 선수 단위 개념 학교 | 초등 |

[표 2. label\_math\_ele 데이터 소개]

본 데이터는 PostgreSQL에 저장된 데이터를 활용하기 위해 sqlalchemy라이브러리를 사용해 데이터베이스 연결을 설정하였다. create\_engine을 통해 데이터베이스 URL을 생성한 후, SQL 쿼리를 사용해 특정 테이블(label\_math\_ele)에서 데이터를 불러와 이를 데이터프레임으로 변환하였다.

**B. education\_2022 데이터 소개**

본 데이터는 label\_math\_ele 데이터에서 from\_id와 to\_id에 존재하는 고유한 knowledgeTag의 id를 가져온 데이터이다. label\_math\_ele에서 ID와 이에 해당하는 from\_chapter\_name 혹은 to\_chapter\_name, from\_semester 혹은 to\_semester를 가지고 왔으며 기존 데이터는 2015 개정 교육과정을 기반으로 이루어졌기에 2022 개정 초등과 수학교육과정을 바탕으로 추가적인 교육과정 태그 및 계열화 작업을 진행하였다. 이 데이터는 6개의 feature와 642개의 sample로 이루어져있다.

|  |  |  |
| --- | --- | --- |
| **feature** | **해석** | **예시** |
| ID | 고유 knowledgeTag id | 2048 |
| Name | from\_chapter\_name 혹은 to\_chapter\_name | 큰 수 > 억과 조를 알아볼까요 / 큰 수 > 억과 조를 알아볼까요 |
| 영역명 | 2022 개정 초등과 수학교육과정 영역명 | 수와 연산 |
| 내용요소 | 2022 개정 초등과 수학교육과정 내용요소 | 다섯 자리 이상의 수 |
| 계열화 | 2022 개정 초등과 수학교육과정 바탕으로 계열화 | A-4-7 |
| grade | from\_semester 혹은 to\_semester에서 학년-학기 추출 | 4-1 |

[표 3. education\_2022 데이터 소개]

다음 2022 개정 수학과 교육과정을 바탕으로 영역명과 내용요소를 설정하여 각 영역명과 내용요소를 바탕으로 계열화(ex. A-1-1)를 진행한 방법이다.

**B-1. 영역별 라벨링**

2022 개정 초등학교 수학과 교육과정은 수와 연산, 변화와 관계, 도형과 측정, 자료와 가능성, 총 4개의 영역으로 이루어져 있으며, 이를 세분화하여 수와 연산, 변화와 관계, 도형과 측정(도형), 도형과 측정(측정), 자료와 가능성으로 나누었다. 각 sample의 Name을 바탕으로 해당 ID의 영역명과 내용요소를 작성하였다.

|  |  |  |
| --- | --- | --- |
| **2022 개정 초등학교 수학과 교육과정 (4개 영역)** | **2022 개정 초등학교 수학과 교육과정 세분화 (5개 영역)** | **영역별 라벨링** |
| 수와 연산 | 수와 연산 | A |
| 변화와 관계 | 변화와 관계 | B |
| 도형과 측정 | 도형과 측정(도형) | C |
| 도형과 측정(측정) | D |
| 자료와 가능성 | 자료와 가능성 | E |

[표 4. 2022 개정 초등학교 수학과 교육과정 영역 계열화]

**B-2. 내용요소 라벨링**

영역별 라벨링을 바탕으로 2022 개정 초등학교 수학과 교육과정의 '2. 내용 체계 및 성취기준' 중 '가. 내용 체계'를 기반으로 영역별 내용요소를 라벨링 하였다. 위에서 아래로, 왼쪽에서 오른쪽을 기준으로 라벨링을 진행하였다.

다음은 수와 연산 영역의 내용요소를 라벨링 한 결과이다. 2015 개정 교육과정에 존재했지만 2022 개정 교육과정에는 삭제된 내용 요소는 17번으로 2022 개정 교육과정 내용요소 이후로 배치하였다. 이로써 내용요소는 17번까지 라벨링이 되었다.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **내용요소** | | | | | | |
| **초등학교** | | | | | | |
| **지식•이해** | **1~2학년** | | **3~4학년** | | **5~6학년** | |
| 네 자리 이하의 수 | 1 | 다섯 자리 이상의 수 | 4 | 약수와 배수 | 11 |
| 두 자리 수 범위의 덧셈과 뺄셈 | 2 | 분수 | 5 | 수의 범위와 올림, 버림, 반올림 | 12 |
| 한자리 수의 곱셈 | 3 | 소수 | 6 | 자연수의 혼합 계산 | 13 |
| 세 자리 수의 덧셈과 뺄셈 | 7 | 분모가 다른 분수의 덧셈과 뺄셈 | 14 |
| 자연수의 곱셈과 나눗셈 | 8 | 분수의 곱셈과 나눗셈 | 15 |
| 분모가 같은 분수의 덧셈과 뺄셈 | 9 | 소수의 곱셈과 나눗셈 | 16 |
| 소수의 덧셈과 뺄셈 | 10 |

※ 분수와 소수의 관계(2015) : 17번

[표 5. 2022 개정 초등학교 수학과 교육과정 수와 연산 내용요소 계열화]

다음은 변화와 관계 영역의 내용요소를 라벨링 한 결과이다. 이로써 내용요소는 6번까지 라벨링이 되었다.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **내용요소** | | | | | | |
| **초등학교** | | | | | | |
| **지식•이해** | **1~2학년** | | **3~4학년** | | **5~6학년** | |
| 규칙 | 1 | 규칙 | 2 | 대응 관계 | 4 |
| 동치 관계 | 3 | 비와 비율 | 5 |
| 비례식과 비례배분 | 6 |

[표. 6 2022 개정 초등학교 수학과 교육과정 변화와 관계 내용요소 계열화]

다음은 도형과 측정 영역의 내용요소를 라벨링 한 결과이다. 도형과 측정 영역은 도형과 측정(도형). 도형과 측정(측정)으로 영역이 세분화 되어 라벨링이 되어 있어 이와 동일한 연장선에서 도형/측정 내용요소를 구별하여 라벨링 하였다. 이 중, 도형과 측정(도형) 영역에서는 2015 개정 교육과정 에는 존재했지만 2022 개정 교육과정에서는 제외된 내용요소를 14번으로 하여, 2022 개정 교육과정 내용요소 이후로 배치하였다. 이로써 도형과 측정(도형)의 내용요소(빨간색)는 14번까지, 도형과 측정(측정)의 내용요소(파란색)는 11번까지 라벨링이 되었다.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **내용요소** | | | | | | |
| **초등학교** | | | | | | |
| **지식•이해** | **1~2학년** | | **3~4학년** | | **5~6학년** | |
| 입체도형의 모양 | 1 | 도형의 기초 | 4 | 합동과 대칭 | 10 |
| 평면도형과 그 구성요소 | 2 | 원의 구성 요소 | 5 | 직육면체와 정육면체 | 11 |
| 양의 비교 | 1 | 여러 가지 삼각형 | 6 | 각기둥과 각뿔 | 12 |
| 시각과 시간(시, 분) | 2 | 여러 가지 사각형 | 7 | 원기둥, 원뿔, 구 | 13 |
| 길이(cm, m) | 3 | 다각형 | 8 | 다각형의 둘레와 넓이 | 9 |
| 평면도형의 이동 | 9 | 원주율과 원의 넓이 | 10 |
| 시각과 시간(초) | 4 | 직육면체와 정육면체의 겉넓이와 부피 | 11 |
| 길이(mm, km) | 5 |
| 들이(L, mL) | 6 |
| 무게(kg, g, t) | 7 |
| 각도( ° ) | 8 |

※ 입체 도형의 공간감각(2015) : 14번

**[표 7. 2022 개정 초등학교 수학과 교육과정 도형과 측정 내용요소 계열화]**

다음은 자료와 가능성 영역의 내용요소를 라벨링 한 결과이다. 2015 개정 교육과정에는 '그림그래프' 내용요소가 3~4학년군과 5~6학년 군 모두에 존재했지만 2022 개정 교육과정에는 5~6학년 군에서 삭제되었기에, 이를 3~4학년 그림그래프와 통합하였다. 이로써 내용요소는 9번까지 라벨링이 되었다.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **내용요소** | | | | | | |
| **초등학교** | | | | | | |
| **지식•이해** | **1~2학년** | | **3~4학년** | | **5~6학년** | |
| 자료와 분류 | 1 | 그림그래프 | 4 | 평균 | 7 |
| 표 | 2 | 막대그래프 | 5 | 띠그래프, 원그래프 | 8 |
| ○, ×, ／를 이용한 그래프 | 3 | 꺾은선그래프 | 6 | 가능성 | 9 |

※ 그림그래프(5~6학년군, 2015) : 4번으로 함께 라벨링

[표 8. 2022 개정 초등학교 수학과 교육과정 자료와 가능성 내용요소 계열화]

**B-3. 내용요소 별 세부적 라벨링**

'영역-내용요소'로 라벨링을 한 후, 내용요소별로 세부적인 라벨링을 진행하였다. 천재교육 1학년부터 6학년의 교과서를 이용하여 내용요소별로 ID의 Name을 참고하여 세부적인 라벨링을 진행하였다. 다음은 예시로 그 중, 도형과 측정(도형) 영역의 네 번째 내용요소의 세부적 라벨링 결과이다.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID** | **Name** | **영역명** | **내용요소** | **계열화** | **grade** |
| 331 | 평면도형 > 선의 종류에는 어떤 것이 있을까요 | 도형과 측정(도형) | 도형의 기초 | C-3-1 | 3-1 |
| 332 | 평면도형 > 각을 알아볼까요 | 도형과 측정(도형) | 도형의 기초 | C-3-2 | 3-1 |
| 2052 | 각도 > 어느 각이 더 클까요 | 도형과 측정(도형) | 도형의 기초 | C-3-3 | 4-2 |
| 2055 | 각도 > 직각보다 작은 각과 직각보다 큰 각을 알아볼까요 / 각도 > 직각보다 작은 각과 직각보다 큰 각을 알아볼까요 | 도형과 측정(도형) | 도형의 기초 | C-3-4 | 4-2 |
| 11235 | 사각형 > 수직을 알아볼까요 / 사각형 > 수직을 알아볼까요 | 도형과 측정(도형) | 도형의 기초 | C-3-5 | 4-2 |
| 11236 | 사각형 > 수직을 알아볼까요 | 도형과 측정(도형) | 도형의 기초 | C-3-6 | 4-2 |
| 11238 | 사각형 > 평행을 알아볼까요 / 사각형 > 평행을 알아볼까요 | 도형과 측정(도형) | 도형의 기초 | C-3-7 | 4-2 |
| 11239 | 사각형 > 평행을 알아볼까요 | 도형과 측정(도형) | 도형의 기초 | C-3-8 | 4-2 |
| 11241 | 사각형 > 평행선 사이의 거리를 알아볼까요 | 도형과 측정(도형) | 도형의 기초 | C-3-9 | 4-2 |

**[표 9. 2022 개정 초등학교 수학과 교육과정 내용요소별 계열화 예시]**

이로써 각 ID에 대해 '영역별(알파벳)-내용요소별(숫자)-내용요소 세분화(숫자)' 형식으로 계열화를 추가하여 개념 간의 관계를 명확히 하였다.

**C. merged\_df 데이터 소개**

본 데이터는 AI Hub의 수학 분야 학습자 역량 측정 데이터 중 성취수준 데이터셋에서 추출한 training 데이터이다. 1학년부터 6학년까지의 데이터를 통합하였으며, 학년별 실력평가에 따른 '1\_문항정오답표', '2\_문항IRT', '3\_응시자IRT' 데이터를 학습자ID(learnerID)를 기준으로 결합하여 구성되었다. 이 데이터는 각 학습자의 문제 정오답, 문항별 IRT(문항반응이론) 지표, 응시자별 IRT 지표를 포함하며, 이를 통해 학습자 성취 수준을 종합적으로 분석할 수 있다. 본 데이터는 7개의 feature와 1677358개의 sample로 이루어져있다. 여기서 7개의 feature에 대한 설명은 다음과 같다.

|  |  |  |
| --- | --- | --- |
| **feature** | **해석** | **예시** |
| learnerID | 학습자 ID  (비식별조치 된 ID로 생성) | A010000003 (1학년 학습자)  A060000009 (6학년 학습자) |
| answerCode | 평가 문항별 정오답 구분 | 0 : 오답, 1 : 정답 |
| knowledgeTag | 수학 지식체계의 노드  고유 아이디(지식체계) | 603, 9232, 8004 등 |
| difficultyLevel | 문항 난이도 | -3.719714746 |
| discriminationLevel | 문항 변별도 | 0.721178751 |
| guessLevel | 문항 추측도 | 0.010089036 |
| theta | 학습자의 이해력 수준 | 0.196765153 |
| 계열화 | 해당 sample의 knowledgeTag 교육과정 계열화 | C-1-5 |

[표 10. merged\_df 데이터 소개]

본 데이터는 MongoDB에 저장된 JSON 파일 형식으로 구성되어 있었으며, 이를 pymongo라이브러리를 사용해 MongoDB에서 불러온 후, 필요한 필드만 선택하여 각 컬렉션을 데이터프레임으로 변환하였다. 이후 학습자 ID(learnerID), 문제 정오답(answerCode), 문항 난이도 및 변별도 등의 정보를 포함한 데이터프레임을 병합하여 분석 가능한 형태로 정리하였다. 추가로, education\_2022 데이터프레임과 merged\_df를 knowledgeTag와 ID를 기준으로 병합하여 학습자의 문제와 관련된 교육 정보(계열화)를 포함시켰다.

**2) 학습 결손 추적 AI 모델 알고리즘 소개**

**A. KST 기반 알고리즘**

본 프로젝트에서 개발한 학습 결손 추적 AI 모델은 [KST 이론을](#kst) 기반으로 개발되었다. 학습자의 결손을 파악하여 그 중 보충 학습 혹은 선행 학습이 필요한 개념들을 선별하여 선수학습과 후속학습을 추천하는 알고리즘을 구성했다.

**B. 선수학습 추천 알고리즘**

다음은 선수학습 추천 알고리즘 순서도이다. 이는 학습자 ID 'A050000008'를 입력하여 출력된 선수학습 추천 알고리즘의 구현 방식을 보여준다.

|  |
| --- |
|  |

**[그림 3. 선수학습 추천 알고리즘 순서도]**

본 AI 모델의 선수학습 추천 알고리즘은 다음과 같다. '1\_문항정오답표', '2\_문항IRT', '3\_응시자IRT' 가 합쳐진 merged\_df를 바탕으로 학습자가 푼 모든 문제들을 뽑아온 뒤, 영역별로(A,B,C,D,E) 존재하는 학습자가 맞힌 문제와 틀린 문제의 각 지식 태그(knowledgeTag)들의 선수학습 개념을 가져와, 맞힌 문제에 포함되지 않은 틀린 문제의 고유한 선수 개념을 추출하여 추천해주도록 구현하였다. 만일 한 영역 내, 틀린 문제의 고유한 선수 개념이 여러개 존재할 경우 딥러닝 모델을 이용하여 학습 결손이 가장 많이 일어난 고유 선수 개념을 한가지 추천하도록 구성하였다.

다음은 선수학습 추천 알고리즘에서 사용한 딥러닝 모델에 대한 소개이다. 본 프로젝트에서 사용한 딥러닝 모델은 학생의 문제 풀이 데이터를 바탕으로 학습 결손을 추적하고, 학습자에게 가장 필요한 맞춤형 개념을 추천하기 위해 설계되었다. 모델을 구축하기 위해 IRT 이론에 기반을 두어 먼저 학생들의 학습 데이터에서 difficultyLevel(난이도), discriminationLevel(변별도), theta(학습자 능력), guessLevel(추측도)의 특징을 추출하고, 학습 목표인 answerCode를 예측할 수 있도록 하였다.

데이터 전처리 과정에서는 계열화 정보를 기준으로 학습을 분류하고, 문자열 형태로 저장된 영역 정보를 숫자로 변환하여 분석의 효율성을 높였다. 단, 변환된 영역 정보는 모델의 입력 특징으로 직접 사용되지는 않았으며, 추천 알고리즘에서 추가적인 분석을 위해 활용되었다.

이 모델은 PyTorch로 구현된 심층 신경망으로, 이진 분류 문제를 해결하기 위해 설계된 EnhancedBinaryNN을 사용하여 학습 결손을 추적하였다. 해당 신경망은 총 5개의 계층으로 이루어져 있으며, 각 계층은 은닉 노드의 수를 점진적으로 줄이며 학습을 하도록 구성되었다.

|  |  |  |
| --- | --- | --- |
| **레이어/연산** | **유형** | **세부 사항** |
| 완전 연결층 (1) | 선형 (Linear) | 입력: 4, 출력: 256 |
| 활성화 함수 (1) | ReLU | ReLU 활성화 함수 적용 |
| 배치 정규화 (1) | BatchNorm1d | 256 유닛에 배치 정규화 적용 |
| 완전 연결층 (2) | 선형 (Linear) | 입력: 256, 출력: 128 |
| 활성화 함수 (2) | ReLU | ReLU 활성화 함수 적용 |
| 드롭아웃 (1) | Dropout | 드롭아웃 확률: 0.4 (40% 적용) |
| 완전 연결층 (3) | 선형 (Linear) | 입력: 128, 출력: 64 |
| 활성화 함수 (3) | ReLU | ReLU 활성화 함수 적용 |
| 완전 연결층 (4) | 선형 (Linear) | 입력: 64, 출력: 32 |
| 활성화 함수 (4) | ReLU | ReLU 활성화 함수 적용 |
| 완전 연결층 (5) | 선형 (Linear) | 입력: 32, 출력: 1 (이진 분류 출력) |
| 출력 활성화 함수 | Sigmoid | Sigmoid 활성화 함수로 최종 출력값 생성 |

[표 11. 딥러닝 모델 은닉층]

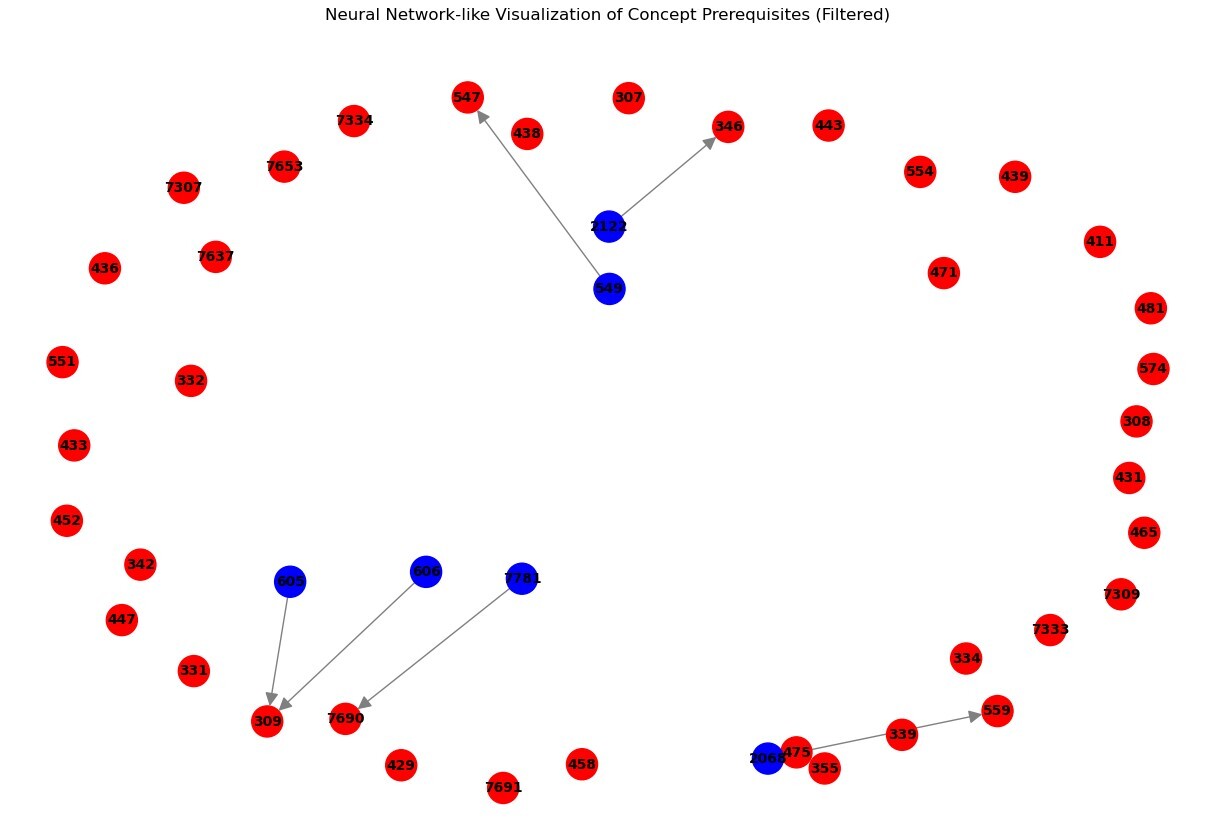
활성화 함수로는 ReLU와 Sigmoid를 사용했으며, 각 은닉층 후에는 배치 정규화(Batch Normalization)를 적용하여 학습의 안정성과 속도를 향상시켰다. 손실 함수로는 이진 교차 엔트로피 손실(BCE)을 사용했고, 옵티마이저는 아담(Adam, 학습률: 0.001)을 사용했다. 하이퍼파라미터로 드롭아웃 비율을 0.4로 설정하고, 모델은 100번의 에포크(epoch) 동안 전체 학습 데이터셋을 사용하여 배치 그라디언트 하강법 방식으로 학습되었다. 매 10 에포크마다 현재 에포크와 손실 값을 출력하여 학습 진행 상태를 추적하였고, 최종적으로 학습된 모델은 'enhanced\_binary\_nn\_model.pth'로 저장되었다.

모델의 성능 평가를 위해 테스트 데이터를 이용해 예측 확률값을 계산하고, 이 값에 기반하여 정확도(Accuracy), 정밀도(Precision), 재현율(Recall), F1 스코어(F1 Score) 등의 성능 지표를 산출하였다.

|  |  |
| --- | --- |
| **Accuracy (정확도)** | 85.56% |
| **Precision (정밀도)** | 0.89 |
| **Recall (재현율)** | 0.89 |
| **F1 Score (F1 스코어)** | 0.89 |

[표 12. 딥러닝 모델의 정확도, 정밀도, 재현율, F1 스코어]

추출된 학습자의 고유한 틀린 문제의 선수학습 개념에 저장된 모델을 사용하여, 예측을 수행하고, predicted\_probability 열에 저장하였다. 'predicted\_probability'는 확률이 높을 수록 해당 개념에 대한 학습 결손이 높다는 것을 의미하며, 추출된 고유한 틀린 문제의 선수학습 개념 중 이 값이 높은 것을 영역별로 한가지 추천하도록 하였다. 다음은 선수 학습 추천 알고리즘을 시각화한 그래프이다.



[그림 4. 본개념과 연관된 선수개념 예시]

위 그래프에서 빨간색 원은 학습자가 틀린 문제의 개념이고, 파란색 원은 틀린 문제의 선수 개념에서 맞춘 문제의 선수개념을 제외한 틀린 문제의 고유한 선수학습 개념을 보여준다. 이 알고리즘에 따라 학습자의 학습 결손을 추적하여 딥러닝 모델을 이용한 맞춤형 선수학습 개념을 추천하도록 구현하였다.

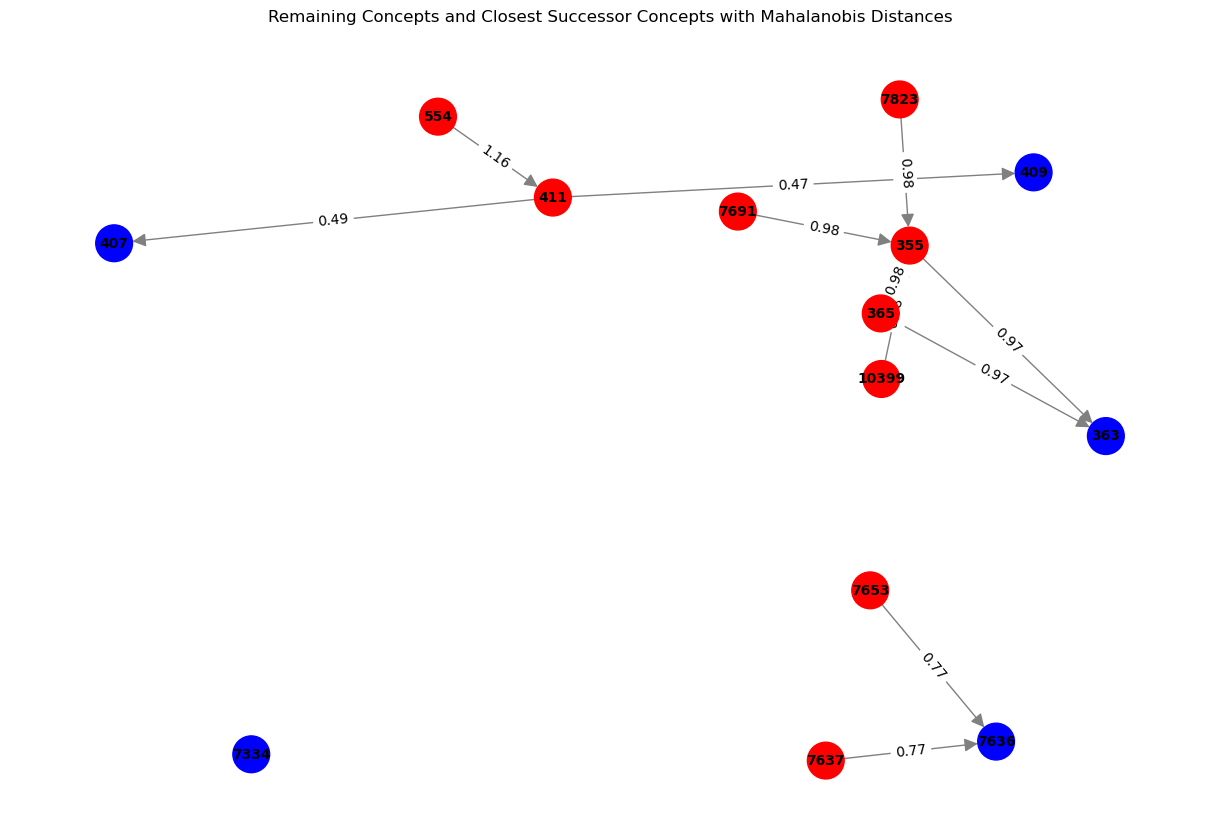
**C. 후속학습 추천 알고리즘**

다음은 후속학습 추천 알고리즘 순서도이다. 이는 학습자 ID 'A050000008'를 입력하여 출력된 후속학습 추천 알고리즘의 구현 방식을 보여준다.

|  |
| --- |
|  |

**[그림 5. 후속학습 추천 알고리즘 순서도]**

본 AI 모델은 KST이론에 기반하여 특정 개념을 잘 이해했을 경우 그와 연관된 후속 개념도 더 잘 이해할 가능성이 높다는 가설에 기반한다. 본 AI 모델의 후속학습 추천 알고리즘은 Manhalanobis 거리와 공분산 행렬을 기반으로 구성하였다. 후속 개념의 특징에 대해 공분산 행렬을 계산하여 학습자가 개념에서 보이는 학습 패턴을 분석하고, 정답 개념의 공분산 행렬과 Manhalanobis 거리가 가장 가까운 후속 개념을 추천한다. 후속학습을 추천하는 경우는 2가지가 있다. 첫째, 맞춘 문제의 선수학습 개념에 틀린 문제의 선수학습 개념이 모두 포함되어 틀린 문제의 고유한 선수학습 개념이 아예 존재하지 않는 경우이다. 둘째, 학습자가 틀린 문제의 고유한 선수학습 개념이 포함된 영역에서는 선수 학습 개념을 추천하고, 만약 이러한 개념이 없는 경우, 해당 영역에서 새로운 후속 학습 개념을 추천하는 경우이다. 예를 들어 학습자가 A,B,C 영역의 문제를 풀었다고 가정했을때, 맞춘 문제의 선수학습 개념과 틀린 문제의 선수학습 개념의 교집합을 제외한 틀린 문제의 고유한 선수학습 개념이 A, B영역일때, C영역은 틀린 문제의 고유한 선수학습 개념에 포함되지 않기에 C영역은 선행학습을 추천해주게 된다. 이 두 경우 모두, Manhalanobis 거리를 이용하여 학습자 맞춤형 후속 학습 개념을 추천하도록 하였다. 다음은 후속 학습 추천 알고리즘을 시각화한 그래프이다.



[그림 6. 거리기반 후속개념 추천 예시]

위 그래프에서 파란색 원은 맞춘 문제의 개념이고, 빨간색 원은 파란색의 원의 후속 개념이다. 빨간색 원들 중 파란색 원과의 거리가 가장 가까운 원이 가장 본개념에 가장 가까운 후속 개념이라 생각하여 이를 추천해주는 후속학습 개념 추천 알고리즘을 구현하였다.

**D. 후속학습 추천 알고리즘 평가 지표**

후속학습 추천 알고리즘의 평가 지표에는 Ranked Precision이 사용되었다. 마할라노비스 거리를 기반으로 후속 학습 개념을 추천한 후, 알맞게 후속 개념이 추천 되었는지 추천된 학습 개념의 마할라노비스 거리 순위를 기반으로 평가한다. 거리가 가까울수록 랭크가 높게 표기 되며, 추천된 학습 개념과 실제로 관련된 개념을 비교하여 Ranked Precision을 계산한다.

**Ⅲ. 결론**

본 프로젝트에서는 학습자의 학습 데이터를 바탕으로 학습 결손을 분석하고, 이를 통해 맞춤형 학습을 제안하는 알고리즘을 구현하였다. MongoDB와 PostgreSQL에 적재된 데이터를 추출해 학습자 ID(learnerID), 문제 정오답(answerCode), 문항 난이도 및 변별도, 학습자의 이해력 등의 정보를 포함한 데이터프레임을 병합하고, 이를 통해 학습자에게 필요한 선수 학습과 후속 학습개념을 추천하는 방식으로 시스템을 구축하였다. 신경망 모델을 활용하여 학습자의 성취도를 예측하고, 맞춤형 학습을 제시함으로써 학습자의 수준에 맞는 후속 학습 개념을 제공하였다.

위 알고리즘을 바탕으로 구현된 본 학습 결손 추적 AI 모델의 결과는 학습자의 학습 결손에 따라 총 2가지 경우로 구현된다.

첫째, 선수 학습과 후속 학습을 모두 추천하는 경우이다. 이는 학습자가 틀린 문제의 고유한 선수학습 개념이 포함된 영역에서는 선수 학습 개념을 추천하고, 만약 이러한 개념이 없는 경우, 해당 영역에서 새로운 후속 학습 개념을 추천한다. 다음은 학습자 A050000008을 해당 학습 결손 추적 AI 모델에 넣어 추출된 결과이다.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **learnerID** | **KnowledgeTag** | **선수학습\_Chapter\_Name** | **계열화** | **영역** |
| A050000008 | 2616 | 자연수의 혼합 계산 > 곱셈과 나눗셈이 섞여 있는 식을 계산해 볼까요 / 자연수의 혼합 계산 > 곱셈과 나눗셈이 섞여 있는 식을 계산해 볼까요 | A-13-2 | 수와 연산 |
| 394 | 다각형의 둘레와 넓이 > 삼각형의 넓이를 구해 볼까요 | D-9-12 | 측정 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **learnerID** | **KnowledgeTag** | **후속학습\_Chapter\_Name** | **계열화** | **영역** |
| A050000008 | 610 | 비와 비율 > 비율을 알아볼까요 / 비와 비율 > 비율을 알아볼까요 | B-5-4 | 변화와  관계 |

[표 13. 선수학습과 후속학습 추천 결과 예시]

A050000008 학습자는 5학년 학생으로 A,B,D 역의 문제를 푼 것으로 확인된다. 학습자가 틀린 문제의 고유한 선수학습 개념이 포함된 영역은 A,D 역이기에 각 영역에서 제시된 선수학습 개념의 'predicted\_probability'의 확률값이 가장 높은 개념은 한 가지씩 추천하였다. B영역은 문제를 풀었지만 틀린 문제의 고유한 선수학습 개념이 포함된 영역이 아니기에, 해당 영역에서 공분산 행렬과 Manhalanobis 거리를 기반으로 가장 가까운 거리에 있는 새로운 후속 학습 개념을 추천한 결과이다.

둘째, 후속학습만 추천되는 경우이다. 학습자가 이미 배운 개념들과 오답을 낸 개념들 간의 선수학습 교집합이 없어 더 이상 학습할 이전 개념이 없다고 판단하여 선수학습을 추천하지 않았을 경우 후속 학습만 추천된다. 다음은 학습자 A040000240을 해당 학습 결손 추적 AI 모델에 넣어 추출된 결과이다.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **learnerID** | **KnowledgeTag** | **후속학습\_Chapter\_Name** | **계열화** | **영역** |
| A040000240 | 710 | 소수의 나눗셈 > (소수)÷(소수)를 알아볼까요 (1) / 소수의 나눗셈 > (소수)÷(소수)를 알아볼까요 (1) | A-16-18 | 수와 연산 |
| 3693 | 규칙과 대응 > 두 양 사이의 관계를 알아볼까요 / 규칙과 대응 > 두 양 사이의 관계를 알아볼까요 | B-4-1 | 변화와 관계 |
| 2123 | 사각형 > 마름모를 알아볼까요 / 사각형 > 마름모를 알아볼까요 | C-6-8 | 도형 |
| 2125 | 꺾은선그래프 > 꺾은선그래프를 알아볼까요 / 꺾은선그래프 > 꺾은선그래프를 알아볼까요 | D-9-2 | 측정 |
| 11265 | 꺾은선그래프 > 꺾은선그래프는 어디에 쓰일까요 | E-6-4 | 자료와 가능성 |

[표 14. 후속학습 추천 결과 예시]

A040000240 학습자는 4학년 학생으로, A, B, C, D, E 모든 영역의 문제를 푼 것으로 확인되었다. 학습자가 이미 배운 개념들과 틀린 문제의 개념들 간에 선수학습 개념의 교집합이 없어 추가로 학습할 이전 개념이 없다고 판단되어, A, B, C, D, E 모든 영역에 대해 Mahalanobis 거리 계산 이론을 기반으로 가장 가까운 후속 학습 개념을 추천하였다.

|  |  |  |  |
| --- | --- | --- | --- |
| **Ranked Precision** | **learner A040000240** | **Area** | |
| Rank 1:  knowledgeTag 2123 | Chapter:  사각형 > 마름모를 알아볼까요 /  사각형 > 마름모를 알아볼까요 | 계열화:  C-6-8 | 영역:  도형 |
| Rank 1:  knowledgeTag 3693 | Chapter:  규칙과 대응 > 두 양 사이의 관계를 알아볼까요 /  규칙과 대응 > 두 양 사이의 관계를 알아 볼까요 | 계열화:  B-4-1 | 영역:  변화와 관계 |
| Rank 1:  knowledgeTag 710 | Chapter:  소수의 나눗셈 > (소수)÷(소수)를 알아볼까요 (1) /  소수의 나눗셈 > (소수)÷(소수)를 알아볼까요 (1) | 계열화:  A-16-18 | 영역:  수와 연산 |
| Rank 1:  knowledgeTag 11265 | Chapter:  꺾은선그래프 > 꺾은선그래프는 어디에 쓰일까요 | 계열화:  E-6-4 | 영역:  자료와 가능성 |
| Rank 1:  knowledgeTag 2125 | Chapter:  꺾은선그래프 > 꺾은선그래프를 알아볼까요 /  꺾은선그래프 > 꺾은선그래프를 알아볼까요 | 계열화:  D-9-2 | 영역:  측정 |

**[표 15. 표 14의 후속학습 추천 결과 평가지표 반영 예시]**

A040000240 학습자의 후속 학습 추천 결과에 대한 평가지표 내용이다. 각 영역별로 랭크가 1로 표기된걸로 보아 각 영역별 추천된 학습 개념과 실제로 관련된 개념이 유사함을 확인할 수 있다.

**Ⅳ. 토론**

본 모델의 선수학습 추천 알고리즘에서 사용한 딥러닝 모델은 AnswerCode(1 또는 0)를 예측한다. 이 결과를 다음과 같이 해석하였다. 모델의 예측 결과가 1(정답)이라면, 입력된 데이터를 통해 학습한 결과가 학습자가 해당 개념에 대해서 잘 알고있기 때문에, 문제를 맞출 확률이 높아 1이라는 값을 도출하였다고 해석하고, 0(오답)이라면 학습자가 해당 개념에 대해 모르기 때문에 문제를 틀릴 확률이 높아 0이라는 값을 도출했다고 해석하였다.

이 해석을 바탕으로 선수학습 추천은 학습자가 이미 학습한 개념을 기반으로 현재 학습 중인 문제와 연관된 이전 개념을 분석하여, 학습 결손을 보완하는 데 중점을 두었다. 선수학습이 존재하지 않는 경우, 학습자가 해당 개념을 충분히 이해하거나 더 이상의 추가 학습이 불필요하다고 판단될 때 후속학습만을 추천하게 된다. 이를 통해 학습자는 자신의 학습 이력을 체계적으로 파악하고 놓친 개념을 보완할 수 있었다.

후속학습 추천은 학습자가 앞으로 학습해야 할 개념을 제안하는 방식으로, Mahalanobis 거리와 공분산 행렬을 사용하여 학습자의 현재 상태와 미래 학습 경로 간의 거리를 효율적으로 계산하여 최적의 후속 학습 개념을 우선순위로 제안하였다. 이를 통해 학습자는 부족한 부분을 보완하면서 동시에 다음 학습 단계를 체계적으로 준비할 수 있도록 하였다.

본 모델은 KST이론과 IRT 기반 문항 분석 지표를 활용하여 학습자의 수준을 정확히 파악하고, Mahalanobis 거리와 공분산 행렬을 통해 학습자의 필요에 맞춘 개별 학습 경로를 제공하는 기능을 구현하였다. 이 시스템은 학습자의 학습 결손을 추적하여 맞춤형 학습 설루션을 제공할 수 있음을 확인하였다. 결론적으로, 본 시스템은 학습자의 성취도를 종합적으로 분석하고, 선수학습과 후속학습을 연계한 개별 맞춤형 학습 경로를 제공함으로써 학습 효율성을 극대화할 수 있었다. 특히, 기초 학력이 부족한 학생들에게 효과적인 학습 경로를 제안할 수 있으며, 교육 현장에서 학습 진단 및 맞춤형 학습 지원에 큰 기여를 할 수 있을 것으로 기대된다.

본 모델은 연구에서 정의된 AI 수학 학습 플랫폼에 필요한 기능 중 수준 분석, 학습 추천, 학습 설계, 학습 이력 데이터 축적 및 분석의 기능을 갖추고 있다.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **특징** | **대기능** | **하위기능** | **개념(설명)** | **서비스대상** |
| 분석 | 학습  분석 | 수준 분석 | - 맞춤형 학습(경로) 및 콘텐츠 등을 제공(추천)하기 위해 학생의 학업성취 수준(시작점)을 진단하고 분석하는 기능 | 학생 |
| 조정 | 학습  추천 | 학습 추천 | - 수준 분석, 실시간 분석, 학습 이력 데이터 분석에 따라 개인별⋅수준별 맞춤형 학습 및 콘텐츠, 문제 등을 제공(추천)하는 기능  - 문제 및 평가 유형: 단원별 보충 및 심화 문제, 차시별 형성평가, 단원별 총괄평가 등  - 오답 피드백: 동일 문제, 유사 문제, 수준 맞춤 문제, 힌트, 관련 동영상, 풀이과정 등 제시 | 학생 |
| 설계 | 교수 설계 | 학습 설계 | - 학생이 스스로의 역량에 따라 학습 경로 및 콘텐츠 구축 | 학생 |
| 보고 | 학습 관리 | 학습 이력  데이터  축적 및 분석 | [인지적]  - 중⋅장기간에 걸친 학습자의 인지적 학습 이력 데이터 축적 및 분석 기능  - 인지적 학습 이력 데이터: 정답률, 오답률, 오답유형, 진도율, 수준 및 성취도, 과제 수행 여부, 문제풀이 시간 등 | 학생,  교사 |

(이기마, 이유정 and 김희정, 2023)[[17]](#footnote-17)

[표 15. 모델의 현재 기능]

이 모델을 추가적으로 발전시킬 수 있는 방향성은 첫 번째로, 학습자의 현재 학습 상태와 새로운 데이터를 즉각적으로 반영할 수 있는 시스템을 구축하면, 더욱 정확한 맞춤형 학습 추천이 가능할 것이다. 두 번째로는 적용 가능성을 확장해 초등학교 수학에서 중학교, 고등학교 수학까지도 적용시킬 수 있을 것이다.

**Ⅴ. 코드 (깃허브) 및 시연 영상**

|  |  |
| --- | --- |
| **Github** | <https://github.com/HKarin426/KST_base_e_math_lack_model.git> |
| **시연영상** | <https://youtu.be/6e-jzhXrm_c?si=2YhPhAK8eqzeGke6> |

**Ⅵ. 출처**

* Doignon, Jean-Paul & Falmagne, Jean-Claude. (2015). Knowledge Spaces and Learning Spaces.
* 박성익, & 임철일. (2022). 교육공학과 수업(9장 테크놀로지 활용 수업의 실제).
* 기초학력 보장법. (2022년 3월 25일 시행) [[URL]](https://www.law.go.kr/lsInfoP.do?lsiSeq=235575#0000)
* Stahl, G., et al. (2022). Knowledge Space Theory [[URL](https://cran.r-project.org/web/packages/kst/vignettes/kst.pdf)]
* Mahalanobis, P. C., “On The Generalized Distance in Statistics,” Proceeding of National Institute of Science of India (Calcutta), Vol. 2, April, 1936, pp. 49~55.
* Jeong, Y., & Sung, Y. (2017). Study on implementation strategies for future digital textbooks. Korea Education & Research Information Service
* Korea Education & Research Information Service. (2021). Status and plans for the implementation of digital textbooks in the 2015 revised curriculum
* Ministry of Education(2023a). Realizing customized education for all: digital-based educational innovation strategies.
* 김순택(1979).교과별 선수능력요인,.서울:한국교육개발원.pp57-70
* 박성익(1985).학습과제의 분석기법,.제38호.pp.13-20
* 정순환. (2015). 분수의 덧셈과 뺄셈에서 부진아의 결손내용 파악을 위한 진단도구 개발 및 적용
* 성지현. (2023). 개인 맞춤형 수학 학습을 위한 인공지능 교육시스템의 기능과 적용 사례 분석. 수학교육, 62(3), 303-326.
* 이기마, 이유정 and 김희정. (2023). 국내 AI 수학 학습 플랫폼의 적응형 학습에 대한 분석. 한국학교수학회논문집, 26(3), 245-268.
* 정제영(2021). 포스트코로나 시대의 미래교육: 비대면지능형 교육 기술의 동향. 융합연구리뷰, 7(3), 4-29.
* 대한민국 청소년활동진흥원, "지능형 튜터링 시스템 - 인공지능 기반 수학교육 학습 플랫폼의 교수학습 전략", 2022-10-04,

**Ⅶ. 라이선스**

|  |  |
| --- | --- |
| **Version** | |
| **python** | 3.12.4 |
| **numpy** | 2.1.1 |
| **pandas** | 2.2.2 |
| **pymongo** | 4.8.0 |
| **scipy** | 1.14.1 |
| **sklearn** | 1.5.1 |
| **sqalchemy** | 2.0.21 |
| **torch** | 2.4.1 + cpu |

|  |  |
| --- | --- |
| **License** | |
| **numpy** | BSD 3-Clause License |
| **pandas** | BSD 3-Clause License |
| **pymongo** | Apache License 2.0 |
| **scipy** | BSD License |
| **scikit-learn (sklearn)** | BSD 3-Clause License |
| **sqlalchemy** | MIT License |
| **torch** | BSD License |

1. Doignon, Jean-Paul & Falmagne, Jean-Claude. (2015). Knowledge Spaces and Learning Spaces. [↑](#footnote-ref-1)
2. 박성익, & 임철일. (2022). 교육공학과 수업(9장 테크놀로지 활용 수업의 실제). [↑](#footnote-ref-2)
3. 기초학력 보장법. (2022년 3월 25일 시행) [[URL]](https://www.law.go.kr/lsInfoP.do?lsiSeq=235575#0000) [↑](#footnote-ref-3)
4. Stahl, G., et al. (2022). Knowledge Space Theory [[URL](https://cran.r-project.org/web/packages/kst/vignettes/kst.pdf)] [↑](#footnote-ref-4)
5. Mahalanobis, P. C., “On The Generalized Distance in Statistics,” Proceeding of National Institute of Science of India (Calcutta), Vol. 2, April, 1936, pp. 49~55. [↑](#footnote-ref-5)
6. 정순환. (2015). 분수의 덧셈과 뺄셈에서 부진아의 결손내용 파악을 위한 진단도구 개발 및 적용 [↑](#footnote-ref-6)
7. 성지현. (2023). 개인 맞춤형 수학 학습을 위한 인공지능 교육시스템의 기능과 적용 사례 분석. 수학교육, 62(3), 303-326. [↑](#footnote-ref-7)
8. 대한민국 청소년활동진흥원, "지능형 튜터링 시스템 - 인공지능 기반 수학교육 학습 플랫폼의 교수학습 전략", 2022-10-04, [↑](#footnote-ref-8)
9. 천재교육 인공지능 취약 진단 학습 추처 시스템 공개특허 공보 (공개일자 2023.5.31, 대한민국특허청) [↑](#footnote-ref-9)
10. J. Doignon & J.-C. Falmagne. Knowledge spaces and learning spaces. In Batchelder et al. (Eds.), New Handbook of Mathematical Psychology (pp. 274-321). Cambridge University Press, 2016. [↑](#footnote-ref-10)
11. Jeong, Y., & Sung, Y. (2017). Study on implementation strategies for future digital textbooks. Korea Education & Research Information Service [↑](#footnote-ref-11)
12. Ministry of Education(2023a). Realizing customized education for all: digital-based educational innovation strategies. [↑](#footnote-ref-12)
13. 정제영(2021). 포스트코로나 시대의 미래교육: 비대면지능형 교육 기술의 동향. 융합연구리뷰, 7(3), 4-29. [↑](#footnote-ref-13)
14. 기초학력 보장법. (2022년 3월 25일 시행) [[URL]](https://www.law.go.kr/lsInfoP.do?lsiSeq=235575#0000) [↑](#footnote-ref-14)
15. 박성익(1985).학습과제의 분석기법,.제38호.pp.13-20 [↑](#footnote-ref-15)
16. 김순택(1979).교과별 선수능력요인,.서울:한국교육개발원.pp57-70 [↑](#footnote-ref-16)
17. 이기마, 이유정 and 김희정. (2023). 국내 AI 수학 학습 플랫폼의 적응형 학습에 대한 분석. 한국학교수학회논문집, 26(3), 245-268. [↑](#footnote-ref-17)