

물질의 상태 분류에 대한 머신러닝 기반 교사교육 프로그램의 교육적 효과

최정인¹ · 백성혜²

¹인천당산초등학교 교사, ²한국교원대학교 화학교육과 교수

목적 본 연구에서는 초·중등 과학교사를 대상으로 물질의 상태 분류에 대한 머신러닝 기반 교사교육 프로그램을 개발 및 실행하여 물질의 상태 분류 머신러닝 모델을 구축하였다. 그리고 모델의 평가 및 과학교사들의 경험을 질적 분석하여 프로그램의 효과성 및 과학수업에 머신러닝을 적용하는 것에 대한 과학교사들의 인식을 확인하고자 하였다.

방법 이를 위하여 중부권 소재의 사범대학의 교육대학원에 재학 중인 초·중등 과학교사 31명을 대상으로 총 3회의 물질의 상태 분류 활동을 수행하고 의사결정트리 알고리즘을 적용한 머신러닝 모델로 구축하였다. 그리고 정확도, F1-score 등 모델 성능 평가를 통해 프로그램의 효과성을 확인하였다. 또한 과학교사들이 머신러닝의 적용에 대해 갖는 생각을 질적 분석하였다.

결과 총 3회의 물질의 상태 분류 활동 동안 구축된 모델의 성능 평가 결과 정확도, F1-score 등의 값이 정적으로 증가하여 물질의 상태 분류에 대한 머신러닝 기반 교사교육 프로그램의 교육적 효과가 있는 것으로 판단되었다. 그리고 연구 참여자들의 머신러닝 모델 생성을 위한 데이터를 분석한 결과, 지식에 근거한 연역적 사고과정을 통해 물질의 상태를 분류하는 경우가 있음이 드러났다. 머신러닝 모델의 분류 성능을 비교한 결과 미시적 분류 기준보다는 거시적 분류 기준이 더 큰 영향을 주었으며, 물질의 사례별 연구 참여자와 머신러닝 모델의 분류 결과를 비교하였을 때 혼합물의 경우 불일치도가 증가하였다. 또한 질적 연구 분석을 통해 연구 참여자들이 머신러닝의 적용에 대해 갖는 생각을 '과학 개념의 생성 과정 체험', '인지갈등', '의사소통' 등 9개의 주제로 도출하였다.

결론 머신러닝을 과학교육의 탐구학습 방안으로 교사교육에 적극 도입하여야 하며, 머신러닝 기반 교육 프로그램을 토대로 귀납적 사고와 같은 과학적 추론 과정을 제대로 반영할 수 있는 교과서 등의 교육 자료가 개발되어야 하겠다. 그리고 혼합물을 물질의 상태 분류 학습에 포함시켜야 하는가에 대한 과학적 관점에서의 논의가 필요하다.

주제어 머신러닝, 인공지능, 물질의 상태, 분류, 과학교사, 교사교육

논문접수: 2022년 10월 31일, 논문심사: 2022년 11월 23일, 게재승인: 2022년 12월 5일

Corresponding to 백성혜, shpaik@knue.ac.kr

이 논문은 2022년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2022S1A5C2A04092614/2022R1A2C2005683).

1. 서론

1. 연구의 필요성 및 목적

우리는 사물인터넷(IoT), 로봇, 인공지능(AI), 빅데이터 등의 기술이 나노기술(NT), 바이오기술(BT), 정보기술(IT), 인지과학(CS)의 융합기술로 발전하고, 이로 인한 사회 구조의 대변혁이 발생하는 4차 산업혁명시대에 살고 있다(WEF, 2016). 다양한 정보통신기술과 스마트 기술이 널리 보급되면서 세상은 빠르게 초연결사

회로 진입하고 있고, 이에 따라 과학교육 환경 역시 큰 변화에 직면하였다. 아울러, 빅데이터에 기반을 둔 인공지능 기술은 교육과 학습에 대한 근본적인 사고 전환을 요구하고 있다. 더 이상 정보와 단순 지식의 습득 만으로는 교육의 가치를 추구할 수 없게 된 것이다. 그러나 지금까지 학교 과학교육은 산업화 시대에 적합한 인재 양성과 기초 교육에 초점을 맞추어 왔다. 이를 개선하기 위해 미래의 과학교육 방향으로 교수학습 측면에서 탐구수행 중심 교수학습, 통합과학적 간학문적 접근, 문제해결 중심의 역량 증진을 강조하고 있으며, 교

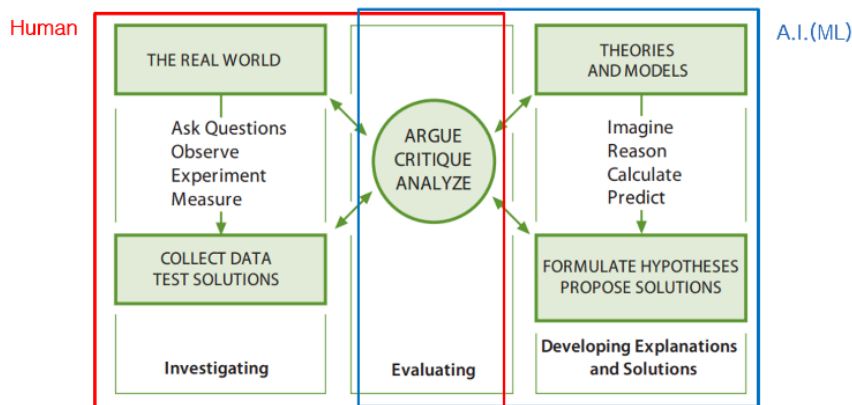
육과정 측면에서는 인공지능 등을 적극 활용하는 교육을 강조하고 있다(송진웅 외, 2019). 지능정보사회를 대비하기 위해 SW 교육, 빅데이터, 인공지능을 가르치는 것은 모든 학생을 데이터엔지니어나 데이터 과학자로 만들려는 것을 목표로 하는 것이 아니다. 이러한 기술은 지능정보사회를 살아가는 데 있어 누구나 갖추어야 할 기본 소양이 될 것으로 보기에 모든 학생을 대상으로 하는 보통 교육의 차원에서 논의되고 있다(가석현, 이민구, 김찬중, 2020; 류혜인, 고아라, 조정원, 2019).

머신러닝(Machine learning)은 과거의 경험을 활용하여 학습하고, 미래의 결정을 내리는 데 그 지식을 사용하는 인공지능의 대표적 하위 분야이다. 그러나 머신러닝은 인공지능의 하위 분야임에도 동의어로 사용되는 경향이 있을 만큼(김진규, 윤길근, 박형근, 2021) 현재 광범위하게 사용되고 있다. 머신러닝은 기계학습으로 번역되기도 하며, 감지 가능한 패턴을 일반화하거나 주어진 예로부터 알려지지 않은 규칙을 만드는 것을 목표로 한다(Dangeti, 2017). 이러한 머신러닝은 레이블(Label)의 유무에 따라 지도학습과 비지도학습으로 구분된다. 지도학습은 가장 일반적으로 사용되고 있는 성공적인 머신러닝 유형으로(Müller & Guido, 2016) 입력 및 출력의 쌍으로 이루어진 데이터를 학습하여 머신러닝 모델을 구축하고, 새로운 입력 데이터에 대해 정확한 예측을 하는 것이 목적이다. 이러한 예측에는 분류(Classification)와 회귀(Regression)라는 두 가지 주요 유형이 있으며, 이 둘을 구분할 수 있는 방법은

레이블의 속성이 연속적인지 이산적인지 살펴보는 것이다. 만약, 물질의 상태 분류와 같은 문제의 경우 출력 값이 고체, 액체, 기체라는 레이블로 연속성이 없는 분류 문제에 해당한다.

이렇게 머신러닝은 데이터를 학습하여 분류 및 회귀의 기능을 자동화된 추론으로 수행할 수 있어 과학교육의 새로운 교육방법이 될 수 있으며, 머신러닝의 도입은 그림 1과 같은 프레임워크(Council, 2012)에서 그 의미를 찾아볼 수 있다.

이 프레임워크는 과학자와 공학자의 활동을 범주로 구분하여 체계화한 것이다. 기존의 프레임워크에 본 연구의 관점을 더하여 인간의 영역과 인공지능(머신러닝)의 영역으로 재해석하였다. 인간의 영역에서 학습자는 의문 생성, 관찰, 실험, 측정 등을 통해 자료 수집의 연구 과정을 수행하게 된다. 그리고 인공지능의 영역에서는 수집된 데이터로부터 추측, 추론, 계산, 예측의 과정이 이루어지게 된다. 프레임워크에서 인공지능 영역의 추측, 추론, 계산, 예측을 위한 머신러닝의 핵심 개념은 알고리즘(Algorithms)이다. 알고리즘은 수학과 컴퓨터 과학, 언어학 또는 관련 분야에서 어떠한 문제를 해결하기 위해 정해진 일련의 절차나 방법을 공식화한 형태로 표현한 것, 또는 계산을 실행하기 위한 단계적 절차를 의미한다. 인공지능 알고리즘이 다른 컴퓨터 프로그램과 구별되는 것은 시각적 인식, 음성 인식, 의사결정, 학습 등과 같이 본질적으로 인간의 영역이라고 생각할 수 있는 분야에 적용될 수 있다는 점이다(김진



[그림 1] 과학자와 엔지니어의 3가지 활동 영역

규, 윤길근, 박형근, 2021). 예를 들어 지도학습의 의사 결정트리 알고리즘의 경우 데이터를 귀납적으로 학습하여 통계에 기초한 자동화된 추론을 수행할 수 있다. 따라서 현행 과학과 교육과정의(교육부, 2015) 내용 체계에서 강조하고 있는 자료의 수집·분석 및 해석, 수학적 사고와 컴퓨터 활용, 모형의 개발과 사용이라는 기능을 고려하였을 때 머신러닝을 과학교육에 활용할 가치가 있다고 볼 수 있다. 그리고 분류를 위한 머신러닝 적용 과학학습은 학습자가 데이터를 생성하는 과정에서 분류를 위한 기준을 생성하고 대상을 관찰하여 분류 기준에 대응하는 속성 값을 입력하여 분류의 결과인 레이블을 부여하는 귀납적 탐구과정을 경험할 수 있다는 장점이 있다. 또한 머신러닝 모델은 새롭게 입력되는 데이터들에 대한 의미 있는 예측치를 제공함으로써 자료와 예측의 일치 정도를 비교·분석하고 비평 및 논증하는 평가의 과정을 가질 수 있도록 도움을 줄 수 있을 것이다. 만약 학습자와 머신러닝 모델의 예측치 사이에 불일치가 발생하게 된다면 이를 해결하기 위해 데이터를 재분석하고, 비판적으로 사고하며, 상호토론을 통한 논증의 과정에서 과학적 개념이 정교화되고, 더불어 과학 학습의 과정 및 과학의 본성에 대한 이해를 촉진할 수 있을 것이기 때문이다.

이러한 측면에서 물질의 상태 분류에 대한 머신러닝 기반 교사교육 프로그램은 과학의 지식, 내용에만 초점을 두는 과학 수업이 아닌 실천(탐구, 기능)을 통해 지식이 발전되어야 함을 전제로 한 것이며, 이러한 실천의 과정에서 과학 지식이 생성되는 것을 이해할 수 있고, 세상을 탐구하고, 모형을 생성하며, 설명을 이끌어내는데 익숙해질 기회를 제공하는 것을 가정하고 있다. 물질의 세 가지 상태 분류는 인공지능을 과학교육에 접목시키기 용이한 주제이며 초·중등 과학 교육과정에서 필수적이고 주요하게 다루어지고 있기에 내실 있는 학습을 위한 인공지능의 효과적 사용을 기대해볼 수 있을 것이다. 특히 물질의 세 가지 상태 분류는 학생 뿐만 아니라 교사들의 사례별 분류 혼란 또한 많으므로(최정인, 백성혜, 2015) 바른 개념 형성을 위해 인공지능을 활용할 수 있다.

현재 우리나라의 인공지능 교육은 정보 교육을 중심으로 다소 편협한 영역에서 진행되고 있는 모습을 볼 수 있다(김영환, 2021). 또한 많은 연구들이 교육의 효과나 학습자 이해도에 중점을 두고 있는데 교사의 이해도나 교사 연수 프로그램 등에 대한 심층적인 연구도 필요한 실정이다(장병철 외, 2021; 한지윤, 신영준, 2020). 그러므로 다양한 교과에서 인공지능 기술의 단순한 적용을 넘어선 심도 깊은 논의를 발전시켜 나갈 필요가 있으며 이를 위해서는 교사를 위한 프로그램의 개발과 적용 역시 요구된다.

과학교육에서 인공지능을 적용한 연구로는 머신러닝을 이용하여 운동과 에너지 단위 학습 프로그램을 제안한 연구(신원섭, 2020), 온라인 학습에서 식물 분류 학습에 적용한 연구(신원섭, 신동훈, 2021), 분자의 구조를 이해하는 융합교육 프로그램을 개발한 연구(이소율, 이영준, 2021), 학생들의 과학 논변 활동의 구성요소 분석의 자동화 가능성을 탐색한 연구(이경건 외, 2018), 화학원소 기호 이미지 학습 프로그램 연구(박민솔 외, 2020) 등이 있다. 과학교육의 다양한 영역에서 인공지능과의 접목이 시도되고 있는 점은 외연 확장의 측면에서 긍정적이라 할 수 있다. 하지만 과학교육에서 머신러닝을 적용한 교사교육 프로그램에 관한 연구는 부족한 상황이며, 프로그램의 효과를 모델의 성능 평가 지표를 통해 확인하는 시도 또한 없는 실정이다. 더불어 물질의 세 가지 상태 분류라는 과학교육의 중요한 학습 주제에 대해 연구 참여자들이 머신러닝 모델의 학습 데이터를 생성하는 과정을 고찰하고, 머신러닝 모델의 분류 결과를 다각도로 비교·분석하는 연구 역시 이루어지지 않았다.

이에 본 연구에서는 중부권 소재의 교육대학원에 재학 중인 현직 초·중등 과학교사를 대상으로 물질의 상태 분류에 대한 머신러닝 기반 교사교육 프로그램을 실행하여 성능 평가 지표를 통해 프로그램의 교육적 효과를 확인하고, 질적 연구 분석 수행을 통해 연구에 참여한 교사들의 기존 교육에 대한 문제점 및 머신러닝 기반 교사교육 프로그램에 대한 인식을 확인하고자 한다. 그리고 연구의 결과를 토대로 머신러닝 기반 교

사교육 프로그램 개발 및 물질의 세 가지 상태 분류 학습에 대한 참고자료를 제공하는데 연구의 목적을 두었다.

2. 연구의 문제

이상의 논의를 바탕으로 본 연구에서는 다음과 같은 연구 문제를 설정하였다.

첫째, 초·중등 과학교사 대상의 물질의 상태 분류에 대한 머신러닝 기반 교사교육 프로그램을 실행하고 성능 평가 지표를 통해 교육적 효과를 확인한다.

둘째, 질적 연구 분석을 통해 기존 교육에 대한 문제점 인식과 머신러닝 기반 교사교육 프로그램에 대한 교사들의 인식을 확인한다.

II. 연구 방법

1. 연구 참여자

연구 참여자의 세부 사항은 표 1과 같다. 이 연구는 중부에 위치한 사범대학의 교육대학원에 재학 중으로 ‘컴퓨팅 사고와 과학교육’이라는 강의를 수강하는 초·중등 과학교사 31명을 대상으로 하였다. 연구 참여자의 일반적 특성은 교육대학원의 전공을 전제로 하고 있으며, 3장의 연구 결과에서 질적 데이터로 제시된 주요 진술을 작성한 초·중등 과학교사들의 경우 교육대학원의 전공별 연구 참여자 코드로 제시하였다. 주요 주제별 의미 있는 진술문을 발췌한 관계로 각 전공의 모든 교사들의 진술이 제시되지 않았다. 한편 모든 연구 참여자들은 머신러닝에 대한 사전 지식과 교육 경험이 없는 상태이다. 연구 참여자의 표본의 수가 적고

유의 표집 방법의 한계가 있으므로 본 연구의 결과를 일반화하는데 있어서 유의하여야 한다.

2. 프로그램 개발 절차

현직 초·중등 과학교사들의 물질의 상태 분류에 대한 머신러닝 기반 교사교육 프로그램을 체계적으로 개발하기 위해 교수설계의 핵심요소인 분석(Analysis), 설계(Design), 개발(Development), 실행(Implementation), 평가(Evaluation)를(강경희, 2013; 박성익 외, 2012) 반영한 교수설계 절차를 따랐다. 구체적인 프로그램의 개발 절차는 다음과 같으며, 개발의 절차 및 세부 내용은 연구자와 화학교육 전문가 1인의 검토를 거쳤다.

분석 단계에서는 초·중등 과학교사들을 위한 물질의 상태 분류에 대한 머신러닝 기반 교사교육 프로그램의 연구 목적을 수립 후 머신러닝에 대한 소양이 있는 교사와 그렇지 않은 교사들을 인터뷰하여 요구 분석 및 학습자 분석을 실시하였다. 설계 단계에서는 물질의 상태 분류에 대한 머신러닝 기반 교사교육 프로그램의 수준과 내용을 정립하고 사용할 머신러닝 프로그램을 선정하였으며 평가 방법을 결정하였다. 실습을 위한 머신러닝 프로그램으로 오렌지(Orange)를 선정하였는데, 오렌지는 오픈소스 머신러닝 툴로 사용 접근성이 좋고, 위젯을 배치 및 연결하는 방식으로 코딩을 쉽게 지원하기 때문이다(Orange, 2021). 개발 단계에서는 인공지능 및 머신러닝, 물질의 상태에 대한 이론, 기초 실습, 적용 실습 등의 내용을 포함하여 물질의 상태 분류 교사교육 프로그램을 개발하였다. 실행 단계에서는 이 연구에서 개발한 교사교육 프로그램을 투입하였으며, 평가 단계에서는 연구 참여자들이 산출한 데이터 및 이를 바탕으로 구축된 머신러닝 모델의 성능 평가 뿐만 아니라 연구에 참여한 소감, 머신러닝 활용 과학

<표 1> 연구 참여자의 세부 사항

교육대학원 전공	인원	질적 분석의 연구 참여자 코드	교육대학원 전공	인원	질적 분석의 연구 참여자 코드
초등과학교육	4	T5, T8, T10, T17	화학교육	9	T1, T2, T3, T11, T16
공통과학교육	6	T4, T6, T15	생물교육	5	T9, T14
물리교육	4	T7, T12, T13, T18	지구과학교육	3	-

교육에 대한 의견을 자유롭게 작성한 에세이를 수집하여 머신러닝 기반 물질의 상태 분류 교사교육 프로그램을 다각도로 검토하였다.

3. 프로그램 개발

프로그램은 인공지능과 머신러닝을 처음 접하는 과학교사의 수준을 고려하여 개발하였다. 초·중등학교에서의 인공지능 교육은 인공지능에 대한 교육과 인공지능 활용교육을 포괄하는 개념으로 정리할 수 있다(홍선주 외, 2020). 인공지능에 대한 교육은 인공지능의 개념과 원리를 이해하고 관련된 기술과 태도를 기르기 위한 교육을 의미하며, 인공지능 활용교육은 인공지능을 다양한 교과와 교수학습 상황에서 교육의 도구로 활용하는 것을 말한다. 인공지능 활용 교육이 성공적으로 수행되기 위해서는 인공지능의 개념과 원리를 이해하는 인공지능에 대한 소양교육이 선행되어야 한다. 따라서 인공지능 및 머신러닝에 대한 이해를 돕는 개관 학습을 선행하도록 하였으며, 물질의 세 가지 상태 분류라는 본 활동에 앞서 척추동물의 분류라는 사전 과제를 수행할 수 있도록 하여 머신러닝에 대한 이해를 돕고 과학교육의 콘텐츠를 머신러닝과 결합하여 문제를 해결할 수 있는 경험의 기회를 제공하도록 하였다. 또한 프로그램은 일련의 주제 중심 프로젝트 및 개인별 과제 수행을 병행하여 자기주도적 교사교육 프로그램이 되도록 하였다.

본 연구에서 개발한 프로그램은 총 9단계로 구성되

었으며, 2주에 걸쳐 교육대학원의 ‘컴퓨팅사고와 과학교육’을 수강하는 초·중등 과학교사들에게 적용되었다. 개발된 물질의 상태 분류에 대한 머신러닝 기반 교사교육 프로그램의 단계별 주요 내용은 표 2와 같다.

교사교육 프로그램은 오리엔테이션 및 인공지능과 머신러닝의 개관(1단계), 머신러닝 지도학습 알고리즘 중 하나인 의사결정트리(Decision Tree) 알고리즘의 이해(2단계), 척추동물 분류에 관한 머신러닝 기초 실습(3단계), 물질의 상태 분류 학습과 관련한 교육과정, 교과서, 선행 연구 고찰(4단계), ML4K를 활용한 머신러닝 기초 실습(5단계), Orange를 사용한 물질의 상태 분류 실습(6~8단계), 머신러닝 기반의 교사교육 프로그램에 참여한 소감 및 머신러닝 활용 과학교육에 대한 의견 작성 시간(9단계)으로 설계하였다.

1단계에서는 인공지능과 머신러닝에 대한 기초 소양을 가질 수 있도록 이론적 개관을 다루게 되며, 2단계에서는 머신러닝의 알고리즘 중 의사결정트리 알고리즘에 대한 이해를 선행하게 된다. 알고리즘은 실제로 직면할 수 있는 다양한 문제를 해결하고자 할 때, 그 문제의 본질을 이해하고 접근할 수 있게 해주기 때문에(정영식, 2015) 적절한 알고리즘의 선정은 중요하다고 할 수 있다. 의사결정트리 알고리즘은 데이터에 있는 규칙을 학습을 통해 자동으로 찾아내 트리 기반의 분류 규칙을 만드는 것이다(권철민, 2020). 본 연구의 내용에 빚대어 설명한다면 연구 참여자들은 물질의 사례(Examples)를 살펴보고 분류 기준(속성, Feature)에

<표 2> 프로그램의 단계별 주요 내용

단계	주요 내용
1	인공지능 및 머신러닝의 개관
2	의사 결정 트리 알고리즘의 이해
3	머신러닝 실습(ML4K를 사용한 척추동물의 분류)
4	물질의 상태 분류 학습과 관련한 교육과정, 교과서, 선행 연구 고찰
5	ML4K를 사용한 머신러닝 기초 실습
6	Orange를 사용한 물질의 상태 분류 실습 프로젝트 Ver.1
7	Orange를 사용한 물질의 상태 분류 실습 프로젝트 Ver.2 - 분류 기준 수정 및 분류 기준 별 속성 값의 입력 내용 수정
8	Orange를 사용한 물질의 상태 분류 실습 프로젝트 Ver.3 - 분류 기준의 합의 및 분류 기준 별 속성 값의 입력 내용 수정
9	프로그램 참여 소감 및 머신러닝 활용 과학교육에 대한 의견 작성

입각하여 원핫인코딩(One-hot encoding)을 한 후 고체, 액체, 기체의 분류 결과(Label)를 쌍으로 입력한다. 원 핫인코딩은 예를 들어 제시된 사례가 분류 기준에 해당하면 1, 해당하지 않으면 0을 입력하는 방식이다. 이렇게 생성된 분류기준 별 속성 값과 분류 결과로 이루어진 데이터 쌍을 머신러닝의 의사결정트리 알고리즘이 학습한 후 새롭게 주어지는 물질을 고체, 액체, 기체 중 하나로 예측 분류하게 되는 것이다.

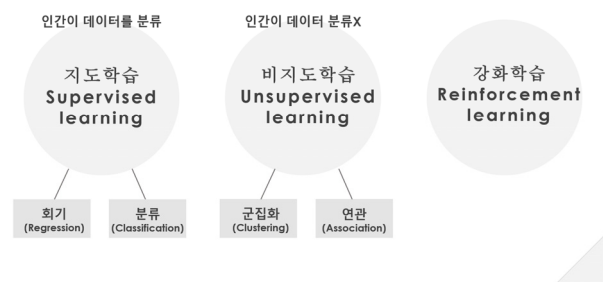
의사결정트리 알고리즘의 경우 가지를 치며 분류를 진행하게 되는데 이때 분류를 위한 기준을 어떻게 선택하는가가 중요한 문제가 된다. 효율적인 노드 분할을 위한 방법으로 정보획득량(Information Gain), 엔트로피(Entropy), 지니(Gini) 계수를 사용하며, 과정의 차이는 있으나 모두 분류 결과의 순도(Purity)를 높이는 방향으로 학습을 진행한다는 공통점이 있다. 이러한 의사결정트리 알고리즘은 쉽고 직관적이며, 비모수 모형으로 의사결정의 과정을 시각화할 수 있다는 특징을 가지고 있어 물질의 상태 분류 학습에 적합하다고 할 수

있다(Dangeti, 2017).

다음의 3단계에서는 머신러닝 학습 플랫폼인 ML4K를 사용하여 척추동물의 계통 분류를 실습하는 시간을 가지도록 하였다. 이를 통해 의사결정트리 알고리즘을 적용한 머신러닝 모델을 구축해봄으로써 머신러닝에 대한 이해를 돕도록 하였다. 그리고 4단계에서는 ‘물질의 상태’에 관한 현행 교육과정 및 교과서, 선행 연구에 대한 내용을 탐색할 수 있게끔 구성하였다.

5단계에서는 ML4K를 사용하여 제시된 사례에 대해 물질의 상태를 분류하는 실습을 진행하도록 하였다. 연구 참여자들은 초등학교 3학년 교과서에 제시된 고체, 액체, 기체의 정의를 기초로 분류 기준을 생성하고, 분류 기준에 해당하면 1, 해당하지 않으면 0을 입력하는 원핫인코딩을 수행하게 된다. 그리고 제시되는 사례의 물질의 상태 분류 결과까지 입력하여 데이터를 생성하고 이를 의사결정트리 알고리즘이 학습하게 한 후 자신의 분류 결과와 머신러닝 모델의 분류 결과를 비교해볼 수 있도록 하였다.

머신러닝의 종류



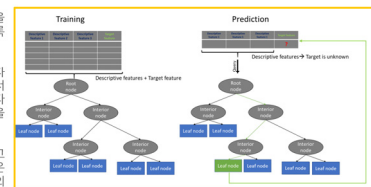
[그림 2] 프로그램 1단계 내용의 예시

의사결정트리란?

각 데이터들이 가진 속성들로부터 패턴을 찾아내서 분류 과제를 수행할 수 있도록 하는 지도학습 머신러닝 모델

-모델의 목표는 입력을 받은 데이터를 보다 작고 동질적인 그룹으로 분할, 여기서 동질적이라 함은 분기별 노드가 보다 순수한(동일 클래스의 비율이 높은) 성격을 갖도록 하는 것을 말함.

- 의사결정나무 모델은 이해도가 높고 직관적이라는 장점이 있으며, 많은 머신러닝 기법과 동일하게 종속변수의 형태에 따라 분류와 회귀 문제로 나뉜다. 범주형 -> 분류, 연속형 -> 회귀



[그림 3] 프로그램 2단계 내용의 예시

다음의 척추동물물 분류 예측하여 봅시다.

데이터	속성A	속성B	속성C	속성D	속성E	속성F	결과 (레이블)
a	X	?
b	?
.	?



[그림 4] 프로그램 3단계 내용의 예시

초등학교 교과서의 물질 세 가지 상태 정의

1. 고체: (나무, 막대, 플라스틱, 막대, 같이) 답는 그릇이 바뀌어도 모양과 부피가 일정한 물질의 상태
2. 액체: (물이나 주스와 같이) 답는 그릇에 따라 모양은 변하지만 부피는 변하지 않는 물질의 상태
3. 기체: (공기, 바람) 답는 그릇에 따라 모양과 부피가 변하고, 담긴 그릇을 항상 가득 채우는 물질의 상태

<2015 개정 교육과정, 국정교과서 초등 과학 3학년>



[그림 5] 프로그램 4단계 내용의 예시

6단계에서는 28가지 물질의 사례를 제시하고 물질의 상태 분류 기준에 입각하여 원핫인코딩을 수행하였다. 그리고 연구 참여자들이 예측하는 물질의 상태 분류 결과까지 스프레드시트 프로그램에 입력 후, 이를 오렌지 프로그램에서 로드하여 의사결정트리 알고리즘으로 머신러닝 모델을 구축하게 된다. 이렇게 구축된 의사결정트리 알고리즘 머신러닝 모델을 Ver. 1로 명명하였다. 6단계에서 선택한 28가지 물질의 사례들은 선행연구를(김선경, 김영미, 백성혜, 2008; 백성혜, 최정인, 박은주, 2013; 유주경, 2012; 최경숙, 2003; 최정인, 백성혜, 2015; 최정인, 백성혜, 2016) 참고하여 사례의 기능을 고려하여 선정하였다.

사례의 기능은 경험적 인지, 속성 파악, 증명에 해당하는 “설명(Explanation)”, 확인 및 비교, 가정에 해당하는 “강화(Reinforcement)”, 그리고 확장 및 대조에 해당하는 “명료화(Clarify)”로 구분할 수 있다(유주경, 2012). 이 중, 설명 및 강화의 기능을 갖는 예들은 보통 개념에 대한 전형성을 지니고 있으며, 같은 예라 하더라도 제시되는 순서, 서술 방식에 따라 사례의 기능이 달라지기 때문에(유주경, 2012) 설명과 강화의 예는 표 3에 제시한 것처럼 한 범주로 통합하였다. 명료화의 사례들은 개념의 속성을 명확하게 하고 개념 속성에 대한 범위를 설명하는 기능을 가지고 있다. 그리고 보통 명료화의 예들은 사례가 제시되는 상황에 의존하는 것이 아닌 사례가 갖는 고유의 특성에 기반하여(유주경, 2012) 확장 및 대조의 기능을 하게 된다. 이와 같은 논의를 바탕으로 설명과 강화의 예로 쇠구슬, 얼음, 드라이아이스, 수은, 질소, 산소, 소금물, 물, 브로민, 이산화탄소, 알코올, 숨, 밀가루, 지우개, 유리, 꿀, 과일주스, 우유, 설탕, 엿, 염소를 선정하였다. 그리고 명료화의 예로 젤리, 쿠키반죽, 구름, 연기, 김(Steam), 안개를 선정하였다.

머신러닝은 데이터에 의존적인 특성이 있다(권철민, 2020). 따라서 편향적이지 않은 데이터를 제공하는 것이 중요하다 할 수 있다. 또한 알고리즘이 데이터를 더 많이 학습할수록 모델이 더욱 정교해진다(Hurwitz & Kirsch, 2018). 이를 위해 연구 참여자 1명당 28개의 분류 결과를 생성하여 전체 연구 참여자가 충분히 많은 데이터를 산출하도록 의도하였다. 분류 활동에 사용되는 사례의 선정은 연구자와 화학교육 전문가 1인 그리고 초등과학교육 박사과정 1인 및 화학교육 박사과정 2인의 검토를 거쳤다.

7단계에서는 이전 단계에서 구축된 머신러닝의 의사결정트리 알고리즘이 예측한 분류 결과와 연구 참여자들의 분류 결과를 비교하고 불일치가 발생하였다면 이를 해소하도록 분류 기준 및 분류 기준 별 속성 값을 수정하는 과정이 진행된다. 이렇게 생성된 데이터로 구축된 의사결정트리 알고리즘 머신러닝 모델을 Ver. 2로 명명하였다.

8단계에서는 연구 참여자의 분류 결과와 의사결정트리 알고리즘을 적용한 머신러닝 모델의 분류 예측 결과 사이의 불일치를 해소하기 위해 다른 연구 참여자들과의 의견 교환을 통해 물질의 상태 분류 기준을 합의하도록 하였다. 상호간 의견의 교환은 전체 인원이 동시에 참여하는 대집단 토론으로 진행하였다. 토론을 통한 학습자 간 상호작용은 자신의 생각을 표출하여 평가받고, 타인과의 타협을 통해 의미 있는 지식을 구성할 수 있게 해주기 때문이다(Richmond & Striley, 1996). 이렇게 분류 기준을 합의한 후 분류 기준 별 속성 값의 입력 내용을 수정하도록 하여 머신러닝 모델이 정교화되도록 하였다. 이 단계에서 구축된 의사결정트리 알고리즘 머신러닝 모델은 Ver. 3으로 명명하였다.

6~8단계의 과정 동안 과학교사들이 어떠한 분류 기준을 고안하고 물질의 상태 분류 활동에 참여하였는지

<표 3> 분류 활동에 사용된 물질의 사례

사례의 기능	사례
설명 및 강화	쇠구슬, 얼음, 드라이아이스, 수은, 질소, 산소, 소금물, 물, 브로민, 이산화탄소, 알코올, 숨, 밀가루, 지우개, 유리, 꿀, 과일주스, 우유, 설탕, 엿, 염소
명료화	젤리, 쿠키 반죽, 구름, 연기, 김, 안개

다음의 물질을 ML for kids로 분류해봅시다.



안개

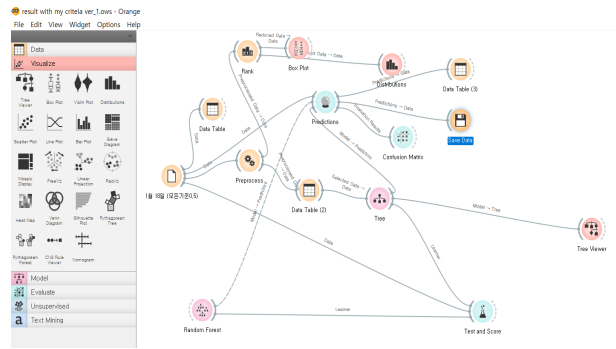


밀가루

[그림 6] 프로그램 5단계 내용의 예시

분류 기준을 거시적 관점과 미시적 관점으로 구분하였다(김선경, 김영미, 백성혜, 2008; 최정인, 백성혜, 2015). 거시적 관점의 분류 기준은 감각적으로 관찰 가능한 특성을 가지며, 미시적 관점의 경우 눈에 보이지 않는 입자의 구조에 대한 것으로 학습에 의해 형성된 것이다.

6단계의 Ver.1 분류 활동에서 연구 참여자들은 거시적 관점의 분류 기준으로 부피 일정, 모양 일정, 공간을 채움, 느낄 수 있음, 흐르는 성질, 단단함, 잡을 수 있음, 합쳤다가 분리 가능, 눈에 보임, 알갱이, 압력에 따른 부피 변화, 덩어리, 점도, 밀도를 제시하였으며, 미시적 관점의 분류 기준으로 입자간 거리, 입자 배열의 규칙성을 제시하였다. 연구 참여자들이 제시한 분류 기준들을 살펴보면 거시적 관점의 분류 기준에 비해 미시적 관점의 분류 기준의 수가 적은 것이 특징이다. 7단계의 Ver. 2 분류활동에서는 연구 참여자의 분류 결과와 머신러닝 모델의 예측 결과와의 불일치를 해소하기 위해 분류 기준을 수정하도록 하였으며, 그 결과 Ver.1의 분류활동에서 제시되었던 분류 기준 가운데 덩어리, 점도, 밀도가 삭제되었다. 그리고 8단계의 Ver. 3 분류 활동에서는 연구 참여자 상호 의견 교환에 의해 분류 기준을 합의하도록 하였는데, 연구 참여자들은 각자 고안한 분류 기준들을 전부 수용하는 것으로 의견을 모았다. 그 결과 거시적 관점의 분류 기준으로 부피 일정, 모양 일정, 공간을 채움, 느낄 수 있음, 흐르는 성질, 단단함, 잡을 수 있음, 합쳤다가 분리 가능, 눈에 보임, 알갱이, 압력에 따른 부피 변화를, 미시적 관점의 분류 기준으로 입자간 거리, 입자 배열의 규칙성으로



[그림 7] 프로그램 6-8단계 내용의 예시

합의하였다.

마지막으로 9단계에서는 물질의 상태 분류에 대한 머신러닝 기반 교과교육 프로그램에 참여한 과학교사들이 교과교육 프로그램에 참여하는 동안 느낀 소감이나 머신러닝에 대한 생각, 의견을 에세이로 작성하는 활동을 포함하였다.

4. 자료 분석

총 31명의 연구 참여자 가운데 연구 참여의 성실성 및 데이터의 품질을 고려하여 Ver. 1, 2 모델의 구축 단계에선 24명, Ver. 3 모델 구축 단계에서는 22명의 연구 참여자 데이터를 자료 분석에 사용하였다. 제시된 물질의 사례들을 분류 기준에 대해 원핫인코딩의 방식으로 0, 1의 속성 값을 입력한 데이터 및 고체, 액체, 기체로 분류한 데이터들이 쌍으로 수집되었다. 데이터 결측치(Missing value)의 경우 평균값(Mean)으로 대체하였으며, 수집된 데이터들은 70:30의 비율로 훈련 데이터와 테스트 데이터로 분할하여 의사결정트리 알고리즘에 학습시키고 예측 모델을 구축하였다. 머신러닝의 모델 성능을 향상시키기 위한 데이터의 전처리 과정에서 속성 값의 결측치가 많을 경우 해당 데이터를 삭제하거나 결측치가 적을 경우 평균값 또는 최빈값으로 대체하는 방법을 사용한다. 본 연구에서는 각 분류 활동 단계에서 발생할 수 있는 어려움을 살펴보는 것이 우선이므로 결측치의 경우 데이터 행(Row)을 삭제하는 대신 평균값으로 대체하는 방법을 선택하였다.

본 연구의 참여자들이 각 사례에 대해 고체, 액체,

기체의 분류 결과를 분류기준 별 원핫인코딩의 입력 내용과 상응하게 부여하였다면, 연구 참여자들의 분류 결과와 머신러닝 모델이 예측한 결과의 일치도가 증가할 것이다. 따라서 이를 확인하기 위하여 각 사례들에 대한 연구 참여자들의 분류 결과와 해당 사례에 대해 머신러닝 모델이 예측한 결과를 혼동 행렬(Confusion matrix)로 나타내어(Dangeti, 2017) 비교하고자 하였다(표 4).

<표 4> 혼동행렬

	예측: Positive	예측: Negative
실제: Positive	TP	FN
실제: Negative	FP	TN

TP(True Positive)는 연구 참여자가 어떤 사례를 액체로 분류하였을 때 머신러닝 알고리즘 모델도 액체로 분류하는 경우를 말한다. TN(True Negative)은 연구 참여자가 어떤 사례를 액체가 아닌 것으로 분류하였을 때 머신러닝 알고리즘 모델도 액체가 아닌 것으로 예측하는 경우를 말한다. 따라서 TP, TN의 빈도가 높을수록 머신러닝 알고리즘 모델의 성능이 우수해진다고 말할 수 있다. FP(False Positive)는 연구 참여자가 고체로 분류하지 않았지만 머신러닝 알고리즘 모델은 고체로 예측한 경우를 말한다. 이는 아무 차이가 없는데도 차이가 있는 것으로 인식시키는 제1종 오류에 해당한다. FN(False Negative)은 어떤 사례에 대해 연구 참여자는 고체로 분류하였지만 머신러닝 알고리즘 모델은 액체 또는 기체로 분류한 경우로 제2종 오류에 해당한다. 위의 혼동 행렬을 토대로 표 5와 같은 머신러닝 알고리즘 모델의 성능지표를 구하여 알고리즘의 성능을 평가할 수 있다(권철민, 2020).

<표 5> 의사결정 트리 모델의 성능 평가 지표

평가지표	공식
정확도	$(TP+TN)/(TP+TN+FP+FN)$
정밀도	$TP/(TP+FP)$
재현율	$TP/(TP+FN)$
F1-score	$2*정밀도*재현율/(정밀도+재현율)$

정확도(Accuracy)는 머신러닝 모델이 전체 데이터 가운데 얼마나 정확하게 예측을 하였는가에 대한 성능 판단 지표이다. 정밀도(Precision)는 머신러닝 모델이 Positive로 예측한 것 중 실제 값이 Positive인 비율을 말한다. 재현율(Recall)은 실제 값이 Positive인 데이터 중에 머신러닝 모델이 Positive로 예측한 데이터의 비율을 뜻한다. F1-Score는 정밀도와 재현율의 조화평균으로 정밀도와 재현율이 어느 한쪽으로 치우치지 않는 수치를 나타낼 때 상대적으로 높은 값을 가지게 된다(권철민, 2020). 연구 참여자들이 사례별 분류 기준의 원핫인코딩 내용이 일관된 경우에 정확도, 정밀도, 재현율, F1-score는 1에 가까워지게 될 것이다. 만약, 1보다 작은 값이 나오면 이는 연구 참여자들이 사례들에 따라 분류 기준의 원핫인코딩 내용에서 일관성을 가지지 못함을 의미한다.

물질의 사례별 분류 결과의 불일치한 정도를 분석하기 위하여 응답 집중 지수(Concentration factor: C-factor)를 사용하였다(Bao & Redish, 2001). 응답 집중 지수는 다음과 같이 정의된다.

$$C = \frac{\sqrt{m}}{\sqrt{m}-1} \times \left(\frac{\sqrt{\sum_1^m n_i^2}}{N} - \frac{1}{\sqrt{m}} \right)$$

m은 문제에서 제시된 답지의 수이며, N은 응답자의 수, ni는 i번째 답지를 선택한 응답자의 수를 말한다. 만약 어떤 사례에 대해 연구 참여자들의 분류 결과가 고체, 액체, 기체 중 특정한 상태에 편중될수록 1에 가까워지며, 분류의 결과가 분산될수록 응답 집중 지수의 값이 작아지게 된다. 한편, 응답 집중 지수를 제안한 Bao & Redish(2001)는 응답 집중 지수의 급간을 정하고 0.5 이상인 경우 응답의 집중 정도가 높다고 하였다.

그리고 연구 참여자들이 작성한 물질의 상태 분류에 대한 머신러닝 기반 교사교육 프로그램에 참여한 소감이나 의견이 들어간 에세이의 자료 분석을 위해 일반적인 질적 연구 분석의 절차를 따랐다. 분석의 단계에는 자료를 코딩하기, 코드를 두드러지는 주제나 패턴을 확인하여 줄이기, 관점 만들어 내기, 그래프나 그림으

로 자료 제시하기의(Madison, 2005) 질적 자료 분석의 핵심요소와 단계를 포함한다. 또한 체계적인 자료 분석을 위해 Taguette 프로그램을(Taguette, 2022) 사용하였다. Taguette는 무료 오픈소스 질적 데이터 분석 소프트웨어(Computer-assisted qualitative data analysis software, CAQDAS)로 질적 데이터의 구조화, 주석 달기, 분석 및 시각화할 수 있도록 도움을 준다(Rampin & Rampin, 2021). 상기의 질적 연구 분석 과정을 반복적으로 수행하고, 코드북을 스프레드시트 파일로 Export한 후 화학교육 전문가 1인과 파일을 공유하였으며, 반복 분석을 통해 코딩의 합의를 이루었다.

III. 연구결과 및 논의

1. 물질의 상태 분류에 대한 머신러닝 기반 교과교육 프로그램의 교육적 효과

과학교사인 연구 참여자들의 물질의 상태 분류 결과와 의사결정트리 알고리즘으로 학습된 머신러닝 모델의 물질의 상태 분류 예측 결과의 일치도를 통해 프로그램의 교육적 효과를 확인하고자 하였다. 이를 위해 머신러닝 알고리즘의 성능 평가 지표인 정확도와 F1-score, 정밀도, 재현율 값을 산출하였다(표 6).

Ver. 1에서 Ver. 3으로 분류 활동이 진행될수록 연구 참여자들의 물질의 상태 분류 결과와 의사결정트리 알고리즘으로 구축된 머신러닝 모델의 분류 예측 결과를 비교하였을 때, 이 모델의 분류 정확도, F1-score, 정밀도, 재현율이 정적으로 증가하는 것으로 나타났다. 이는 과학교사들이 제시된 물질들에 대해 분류 기준별 원핫인코딩의 값을 입력하는 과정이 점점 정교화 되면서 분류의 결과인 레이블이 일치해가는 과정을 나타내

며 그에 따라 머신러닝 모델의 분류 예측 결과 역시 증가함을 의미한다. 따라서 교과교육 프로그램의 교육적 효과가 있는 것으로 판단된다.

각 평가 지표가 1에 가까울수록 머신러닝의 의사결정트리 알고리즘 모델의 일반화도가 높다고 말할 수 있는데 모든 분류 활동에서 각 평가지표는 1보다 작은 값을 나타내고 있었다. 이것은 연구 참여자들이 각 물질들의 사례에 대해 물질의 상태를 다르게 판별하였다는 것을 말하며, 또는 같은 물질의 상태로 판별하였다는 하더라도 분류 기준에 대한 원핫인코딩의 값을 다르게 입력하여 머신러닝의 의사결정트리 알고리즘의 학습 결과가 달라졌음을 말한다. 이러한 차이가 발생한 까닭을 살펴보기 위해 Ver. 3의 분류활동에서 연구 참여자들이 입력한 원핫인코딩의 내용을 일부 발췌하였다(표 7).

교사 A는 연기가 부피와 모양이 일정하고, 공간을 채울 수 없으며, 느낄 수 있고, 흐르는 성질이 없으며, 단단하고 잡을 수 있으며, 합쳤다가 분리가 불가능하고 눈에 보이며, 입자간 거리가 가깝고, 가루 물질 형태이며, 압력에 따른 부피 변화가 없고 분자 배열이 규칙적이라고 응답하였다. 그리고 연기를 고체로 예측하였다. 머신러닝 모델의 경우도 0.977의 확률로 연기를 교사 A와 마찬가지로 고체로 분류하였다.

머신러닝의 의사결정트리 알고리즘은 훈련 데이터들에 대해 엔트로피(Entropy)를 낮추고 순도(Purity)를 높이는 방향으로 학습을 진행하게 된다. 다시 말하면 각각의 분류 기준이 모델 생성에 차등적으로 도움을 줌을 의미한다. 이와 관련하여 어떤 분류 기준이 규칙 트리를 만드는데 기여하였는지를 나타내는 특성 중요도(Feature importance)를 표 8과 같이 출력할 수 있다(Orange, 2021). 표 8의 특성 중요도에 비추었을 때 교사 A는 상위 중요도를 갖는 분류 기준의 속성 값이 모두 고체의 속성을 나타내는 응답을 하였다. 교사 A가

<표 6> 물질의 상태 분류에 대한 의사 결정 트리 모델의 성능 평가 결과

	정확도	F1-score	정밀도	재현율
Ver. 1	0.820	0.820	0.826	0.820
Ver. 2	0.825	0.823	0.824	0.825
Ver. 3	0.909	0.908	0.908	0.909

<표 7> Ver. 3의 분류 활동에서 연기 사례에 대한 연구 참여자와 머신러닝 모델의 예측 세부 내용

연구 참여자		A	B
분류 기준	일정한 부피	1	0
	일정한 모양	1	0
	공간 채움	0	1
	느낄 수 있음	1	1
	흐르는 성질	0	0
	단단함	1	0
	잡을 수 있음	1	0
	합쳤다가 분리 가능	0	1
	눈에 보임	1	1
	입자간 거리	1	0
	가루 물질	1	0
	압력에 따른 부피 변화	0	1
	분자 배열의 규칙성	1	0
연구 참여자의 예측		고체	고체
머신러닝 모델의 예측		고체	기체
머신러닝 모델의 분류 확률	고체	0.977	0.200
	액체	0.023	0.000
	기체	0.000	0.800

<표 8> Ver. 3 분류 활동의 분류 기준별 특성 중요도 순위

순위	분류 기준	특성 중요도 (평균±표준편차)	순위	분류 기준	특성 중요도 (평균±표준편차)
1	부피 일정	0.160718 ± 0.0082007	8	합쳤다가 분리 가능	0.028513 ± 0.0046024
2	잡을 수 있음	0.109562 ± 0.0104094	9	분자 배열의 규칙성	0.023484 ± 0.0032918
3	흐르는 성질	0.081490 ± 0.0070283	10	가루 물질	0.023358 ± 0.0055453
4	단단함	0.063949 ± 0.0041910	11	공간 채움	0.020211 ± 0.0044097
5	압력에 따른 부피 변화	0.039460 ± 0.0073446	12	눈에 보임	0.019367 ± 0.0023217
6	모양 일정	0.034280 ± 0.0059883	13	느낄 수 있음	0.017458 ± 0.0016503
7	입자 간 거리	0.028855 ± 0.0064373	-	-	-

연기를 고체로 분류한 것은 연기를 구성하는 작은 고체 미립자에 초점을 둔 것이며, 이를 토대로 원한인코딩을 수행하여 교사 A와 머신러닝 모델 모두 연기를 고체로 분류하였다.

반면, 교사 B는 연기를 고체로 분류하였지만 머신러닝 모델의 경우 해당 물질을 기체로 예측하였다. 표 8의 상위의 특성 중요도를 갖는 분류 기준의 속성에 대해 부피가 일정하지 않고, 잡을 수 없으며, 흐르는 성질이 없고, 단단하지 않으며, 압력에 따른 부피 변화가 있고, 모양이 일정하지 않다고 응답하여 머신러닝 모델의 경우 0.800의 확률로 연기를 기체로 분류하게 된 것이다. 즉 교사 B는 연기가 고체라는 지식을 가지고 있었지만 분류 기준에 대한 속성 값은 일관되지 않게

입력했다고 볼 수 있다. 이러한 교사 B의 사례는 일상의 경험이 수업이나 교과서에 제시된 지식과 상충함으로써 기인한다는 주장(Stepans, 1989)과 연관된다.

한편 미시적 분류 기준이 28가지 물질의 사례들에 대한 데이터들로 학습하여 구축된 머신러닝 의사결정 트리 알고리즘 모델의 분류 성능에 영향을 끼쳤는지 알아보기 위하여 미시적 분류 기준에 해당하는 ‘입자 사이의 거리’와 ‘분자 배열의 규칙성’의 원한인코딩 값을 제외시키고 학습을 시킨 후 모델을 구축하였다. 그 결과는 표 9와 같다.

미시적 분류 기준이 포함된 경우(표 6)와 포함되지 않은 경우(표 9)를 비교하였을 때 Ver. 1의 분류 활동에서 생성된 모델의 경우 정확도 및 F1-score, 정밀도, 재

<표 9> 물질의 상태 분류에 대한 의사 결정 트리 모델의 성능 평가 결과(미시적 관점의 분류 기준 제외)

	정확도	F1-score	정밀도	재현율
Ver. 1	0.820	0.820	0.826	0.820
Ver. 2	0.795	0.794	0.797	0.795
Ver. 3	0.911	0.911	0.911	0.911

현을 모두 동일하게 나타나 영향을 미치지 않는 것으로 나타났다. Ver. 2의 분류 활동의 경우 미시적 관점의 분류 기준을 제외시켰을 때 각 성능 평가 지표가 미시적 관점의 분류 기준을 포함시켰을 때 보다 하락한 것으로 나타났다. 반면, 연구 참여자들 간 분류 기준에 대한 합의가 이루어진 후 실시된 Ver. 3의 분류 활동 후 생성된 모델의 경우 미시적 분류 기준을 포함시키지 않았더라도 각 성능 평가 지표가 소폭 상승하였다. Ver. 3의 머신러닝 모델의 경우 분류 활동 경험의 축적 및 합의의 과정을 통해 가장 정교화 된 상태라고 볼 수 있다. 따라서 본 연구에서는 미시적 분류 기준의 유무가 모델의 분류 예측 결과에 큰 영향을 주지 않은 것으로 나타났다. 이것은 표 8에서 살펴본 바와 같이 Ver. 3의 분류 활동에서 상위의 특성중요도를 가진 분류 기준들이 거시적 관점의 분류 기준이었다는 결과와 상응한다.

물질의 상태 분류에 대한 의사결정트리 모델을 생성함에 있어서 알고리즘은 입력된 데이터들을 귀납적으로 학습하여, 통계를 기반으로 모델을 구축하게 된다. 위의 결과를 통해 물질의 상태 분류에 있어 거시적 분류 기준의 적용만으로도 분류 결과를 도출할 수 있으며, 미시적 관점의 분류 기준을 적용하는 것이 실효성이 있는가에 대한 의문을 제기할 수 있다. 더욱이 귀납적 사고의 과정은 관찰이나 실험으로부터 데이터를 수집하여 귀납 추론을 통해 경험적인 일반화를 얻는 것

이라고 보았을 때 직접적인 데이터 수집이 어려운 미시적 분류 기준의 적용은 분류를 위한 증거 생성이 아닌 학습된 개념을 적용하고 현상을 해석하는 연역적 사고과정이 포함되어 있다고 보아야 할 것이다. 즉, 추론된 일반화로부터 연역적으로 사고하여 현상을 해석하고 설명함으로써 결과를 정당화하는 기능을 수행하는 것이다.

머신러닝의 의사결정트리 알고리즘 모델이 예측한 분류 결과와 연구 참여자들이 분류한 결과가 고체, 액체, 기체별로 어느 정도 일치하는지를 알아보기 위해 혼동 행렬을 출력하였다(그림 8).

제시된 혼동 행렬과 같이 머신러닝의 의사결정트리 알고리즘이 예측을 Positive로 한 대상 중 연구 참여자들의 분류 결과가 Positive로 일치한 데이터의 비율을 비교한 결과 분류 활동이 진행될수록 정밀도가 증가하였다.

물질의 28가지 사례 별 분류 결과의 불일치한 정도를 분석하기 위하여 응답 집중 지수(C-factor)를 구하여 활동별 응답의 분산 정도가 큰 사례 6가지를 제시하였다(표 10). 응답 집중 지수는 선다형 문항에서 응답별 집중 정도를 나타낸다. 세 가지 물질의 상태 중 하나의 상태로 분류 결과가 모아지면 1에 가까워지며, 물질의 상태 분류 결과가 분산될수록 응답 집중 지수의 크기는 작아지게 된다. 대체적으로 연구 참여자의 분류 결과의 응답 집중 지수와 머신러닝 모델의 분류 예측 결

Ver. 1						Ver. 2						Ver. 3					
Actual	Predicted					Actual	Predicted					Actual	Predicted				
		고체	기체	액체	Σ			고체	기체	액체	Σ			고체	기체	액체	Σ
	고체	91.8 %	1.2 %	8.3 %	259		고체	96.0 %	0.0 %	3.7 %	247		고체	97.9 %	0.8 %	3.7 %	247
	기체	3.5 %	84.5 %	6.7 %	162		기체	0.0 %	99.2 %	3.7 %	130		기체	0.0 %	99.2 %	0.4 %	130
액체	4.7 %	14.3 %	85.0 %	251	액체	4.0 %	0.8 %	92.7 %	239	액체	2.1 %	0.0 %	95.9 %	239			
Σ	257	161	254	672	Σ	248	122	246	616	Σ	242	130	244	616			

[그림 8] 물질의 상태 분류 결과에 대한 연구 참여자와 머신러닝 모델 간의 혼동 행렬

<표 10> 물질의 주요 사례별 연구 참여자와 머신러닝 모델의 분류 결과의 응답 집중 지수 비교

사례	Ver. 1 응답 집중 지수		Ver. 2 응답 집중 지수		Ver. 3 응답 집중 지수	
	연구 참여자의 예측	머신러닝 모델의 예측	연구 참여자의 예측	머신러닝 모델의 예측	연구 참여자의 예측	머신러닝 모델의 예측
연기	0.27770	0.08286	0.27770	0.24184	0.23638	0.03621
김	0.25388	0.20515	0.25388	0.12908	0.51527	0.50294
브로민	0.35850	0.41396	0.35850	0.33008	0.59946	0.51527
구름	0.39199	0.14203	0.39199	0.10945	0.51527	0.42067
쿠키 반죽	0.39751	0.33008	0.39751	0.39199	0.47173	0.33444
안개	0.55067	0.25388	0.55067	0.18648	0.79563	0.70267

과의 응답 집중 지수는 정적인 상관을 보이고 있으며, 혼합물의 경우 응답의 집중 정도가 낮은 것으로 나타났다. 분산매인 공기 중에 고체나 액체 분산질이 부유하고 있는 연기, 김, 구름 등이 그에 해당한다. 물질을 상태 분류 할 때 그것을 구성하고 있는 성분에 초점을 두느냐 또는 성분이 포함된 전체에 초점을 두느냐 상태 분류의 결과가 달라지기 때문에 응답 집중 지수 역시 낮게 나타나게 되었다. 비교적 적은 양의 물에 밀가루가 섞여있는 쿠키반죽도 물과의 혼합 비율에 따라 점성과 탄성을 동시에 나타내어 분류를 어렵게 한다.

분류 활동이 진행될수록 응답 집중 경향을 높아져 Ver. 3의 분류활동에서는 연구 참여자의 분류 결과와 머신러닝 모델의 분류 예측을 비교하였을 때 연기와 쿠키 반죽을 제외한 나머지 사례들에 대해서는 어느 정도 응답 집중이 이루어진 것으로 보인다. 이는 Ver. 3의 분류 활동에서는 분류 기준에 대한 상호 합의가 이루어져 모델이 학습하여야 할 데이터의 품질이 좋아졌기 때문이다.

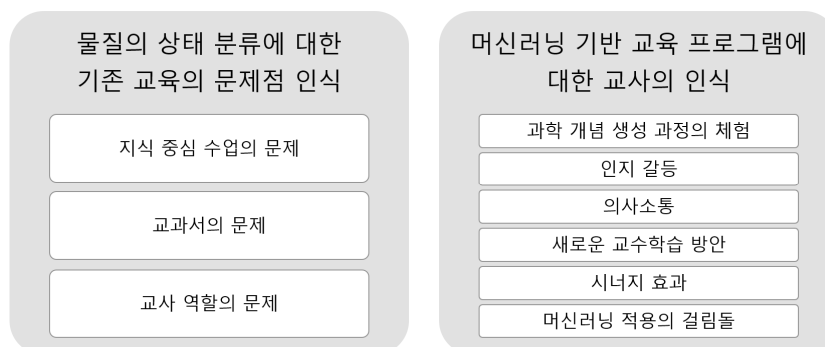
2. 기존 교육에 대한 문제점 및 머신러닝 기반 교육 프로그램에 대한 인식

물질의 상태 분류에 대한 머신러닝 기반 교육 프로그램에 참여한 과학교사들의 소감 및 기존 교육에 대한 문제점 인식이 담긴 에세이를 질적 연구 소프트웨어인 Taguette를 이용하여 분석하였다. 주요 진술문을 하이라이트 처리하고, 하이라이트 처리된 진술들을 수평화하면서 의미 단위를 코드로 태깅하는 방식으로 진행하였다. 이후 코드북을 Export하여 화학교육전문가 1인과 공유 및 교차 반복 분석을 수행하면서 두드러지는 주제로 묶어가면서 코딩을 합의하였다. 그 결과 93개의 의미 있는 진술에 기반한 9개의 주요 주제가 나타났다(그림 9).

가. 물질의 상태 분류에 대한 기존 교육의 문제점 인식

1) 지식 중심 수업의 문제

연구 참여자들은 물질의 상태 분류에 대한 머신러닝 기반 교육 프로그램에 참여하는 동안 과학교사로서 현



[그림 9] 기존 교육에 대한 문제점 및 머신러닝 기반 교육 프로그램에 대한 인식

재의 과학 수업이 지식 중심으로 흘러가는 것에 대한 문제의식을 공유하고 있었다. 이 주제에는 과학적 사고 탐구 과정이 부족한 수업 성찰, 연역적 사고 과정을 통한 분류, 정답과 암기, 진도와 평가에 대한 부담, 교육 과정의 한계 등의 코드가 포함된다.

T6교사: 물질별 속성을 생각할 때는 0 같은데, 물질을 보니 '1이 되어야 하는데' 하는 생각에 약간의 조작이 있었던 것 같다. (몇 개 되지는 않지만, 결정적 속성이었을 것 같다) 물질의 속성을 있는 그대로 보지 않고, 인정하지 않고, 연역적으로 입력하는 모습을 보고...

T6교사는 분류 기준에 대한 원핫인코딩을 수행하면서 속성을 있는 그대로 입력하는 것이 아닌 물질의 상태 분류 예측 결과에 맞추어 속성 값을 부여하였다고 하였다. 그리고 그 과정에서 스스로 연역적으로 사고하였음을 말하고 있다. 분류의 경우 개별적인 관찰 대상들에서 공통성을 찾아 집단으로 구분하고 다른 대상들과의 차이점을 찾는 전략이다. 이러한 분류는 개별적인 사물에서 공통 특징을 찾아서 군집화하여 새로운 대상에 일반화하는 대표적인 귀납 추론의 유형으로 알려져 있다(성현란, 1989). 귀납과 연역은 흐름이 반대인 사고방식으로(이선경, 2015) 분류에서 요구하는 귀납적 사고가 아닌 연역적 사고가 수행되었다는 것은 사례들을 통해 물질의 상태에 대한 개념을 구축, 확장한 것이 아니라 개념을 개별 사례에 적용하여 확인하는 과정이 이루어진 것으로 볼 수 있다. 즉, 관찰 결과를 토대로 분류가 이루어지는 것이 아니며, 이미 물질의 상태 분류 결과를 정하여 놓고 그것에 맞추어 분류 기준에 대한 속성 값을 입력한 것을 의미한다. 이것은 이미 정해진 답에 끼워 맞추는 방식으로 수행되기에 새로운 지식으로의 확장을 불가능하게 한다. 이러한 연역적 추론은 설명식 교수학습의 핵심적 과정으로 알려져 있다. 연역적 교수학습은 과학적 개념, 법칙, 원리, 이론 등을 정의하거나 설명하고, 그 실례를 도출하여 상세하게 기

술하거나 확인하는 등의 과정을 거쳐 수행된다(조희형 외, 2014).

T15교사: 항상 문제 풀이를 하면서도 '기본 정의가 중요하니 반드시 외워!'라고만 했지 '이 정의가 왜 이렇게 나왔는지 우리 한번 얘기 해 보자!'고 한 적이 거의 없었던 것 같습니다.

T3교사: 기준이 주어지고 변하면 평가에서 틀릴 수도 있기 때문에 학생들이 암기식으로 학습하게 됩니다.

T15교사는 교사로서 정해진 답의 전달 및 암기에 치중한 본인의 수업에 대한 성찰을 하였다. 그리고 개념의 정의가 나온 그 과정에 대한 논의 즉, 과학적 사고 과정이 빠진 수업에 대한 이야기도 하고 있다. T3교사가 말한 것처럼 학생 평가라는 현실적인 제약을 고려해야겠으나 위에서 기술한 바와 같이 교사들이 갖는 연역적 사고 경향도 정답과 암기 위주의 수업으로 흘러가는데 영향을 주었는지 살펴볼 필요가 있다. 교사가 학생들에게 과학을 가르칠 때 교사 자신이 가지고 있는 과학에 대한 인식도 학생들에게 함께 전달되며 이러한 철학적 부분은 과학에 대한 이미지 형성에 영향을 미친다는 면에서(Matthews, 1994, 권성기, 송진웅, 박종원 역, 2014) 과학 지식은 불변의 진리라는 인식을 학생들이 갖게 될 수 있다. 또한 암기한 지식의 양보다는 지식의 질이 중요하며, 어떤 정보가 어디에 있는지, 어떻게 찾을 수 있는지를 아는 것이 필요하다. 즉, 지식 축적보다는 지식 판별과 활용능력이 중요해지는 시대의 흐름을(최연구, 2017) 과학교육에서도 간과해서는 안 될 것이다.

T1교사: 현재 교육과정은 나선형으로 학교급이 높아질수록 같은 내용 요소가 점차 심화되도록 구성되어 있다. 하지만 실질적으로 심화된 것이 아니라 분절된 것처럼 느껴질 때가 있었다.

T1교사는 현 교육과정 구성의 문제에 대해 이야기를 하였다. 특히 초등학교에서 중학교로 학교급이 높아지면서 내용이 연속적으로 심화되는 것이 아닌 분절적이라고 느끼고 있었다.

2) 교과서 문제

연구 참여자들은 지식 중심의 수업 문제와 함께 교과서의 문제도 인식하고 있었다. 교과서 문제라는 주제는 현 교과서 구성에 대한 불만족, 안내된 탐구 따라하기, 교과서 중심의 수업에 대한 성찰, 전형적인 물질 사례만을 제시한 교과서, 학습량 과다 등의 코드를 포함한다.

T9교사: 교과서 구성이 이렇게 되면 학생들의 입장에서 원핵생물계에서 동물계에 이르기까지의 내용을 암기할 수밖에 없을 것 같다. 교과서 흐름을 보면 100쪽, 101쪽에 원핵생물계, 원생생물계, 식물계, 균계, 동물계의 특징이 먼저 나오고 102쪽에서 주어진 생물을 각 계의 특징에 맞게 분류하도록 되어있다. 이러한 교과서 흐름은 각 계의 특징을 무조건 암기하도록 구성되어 있다고 생각한다.

T10교사: 과학 개념을 설명하기 위해 그 개념에 도달할 수 있는 정해진 실험을 소개하고 그 실험을 통해 밝혀진 개념을 배운 활동이 반복되어 있었다. 실험도 조건과 방법이 다 안내되어 있었고, 아이들은 그 과정을 기계처럼 실시하고 사실을 단순히 확인하는 활동의 반복이었다. 학생들 간의 학습 격차도 크고 가지고 있는 기본 지식도 다 다르기 때문에 이러한 방법이 효율적일 수 있지만, 과연 최선의 방법일까 의문이 들었다. 현재 과학 교육에서 강조하고 있는 탐구력을 교과서에 있는 실험과 수업을 학습하여 단순히 관찰하고 확인하는 반복 활동을 통해 이끌어 낼 수 있을지에 대해 의문이 생겼다.

생물교육 전공의 T9교사는 중학교 1학년 교과서의 생물 사례를 들며 각 생물계의 개념이 먼저 제시되고 이후 제시된 생물 사례를 각 계에 맞게 분류하는 활동으로 구성되어 있다고 소개하며 이러한 교과서의 구성은 학생들의 암기를 유도하는 하는 구성이라고 지적하고 있다. 이를 통해 물질의 상태 분류라는 화학 사례를 생물 사례에 적용하여 교과서의 문제를 인식하고 있음을 알 수 있다. 또한 T10교사는 안내된 실험을 따라하는 교과서 구성의 효율성을 인정하면서도 이러한 교과서의 구성에서 학생들의 탐구력을 향상시킬 수 있을지에 대한 의문을 표하고 있다. 이는 분류능력을 배양시키기보다는 타인의 두뇌(학자나 교사)를 통해 이미 분류된 후 교과서나 참고서에 제시된 결과적 분류 지식을 암기하는 방식이 주를 이루고 있다는 선행연구와도 일치하는 의견이다(권용주 외, 2011).

T4교사: 교과서의 깔끔하게 정리되어 있는 명확한 사례만을 가지고 연역적 사고를 토대로 오류 없이 가르치는 것을 넘어서서 애매 모호한 다양한 사례에 노출시켜...

T5교사: 초등학교 3학년 교과서도 고체, 액체, 기체를 분류하는 애매한 기준은 빼고 가르치고 싶은지 고체, 액체, 기체를 같은 분류기준으로 분류하는 과정이 없고 고체, 액체, 기체 상태의 대표적인 물체 하나를 관찰하고 그 특징을 찾아보는 활동으로 구성되어 있었다.

T4교사와 T5교사는 모두 교과서에 학습하여야 할 개념의 전형성을 띤 사례만이 제시되어 있음을 언급하며, 분류기준이 명확한 사례만을 제시하는 것이 아닌 애매모호한 다양한 사례를 제시할 필요성을 말하고 있다. 수업은 교과서의 내용에서 벗어나기 어려운 것이 현실이므로 교수학습의 주 자료로서 교과서는 매우 중요한 위치를 차지하고 있다. 교과서는 수업의 목표와 주요 개념, 그리고 주요 개념에 대한 설명 및 관련 개념과 사례 등이 조직되어 있는 수업 자료이다(이선경,

2012). 만약 분류라는 탐구 과정을 통해 획득되어야 할 개념이 있다면 교과서 역시 분류를 통한 귀납적 개념 획득이 이루어지도록 구성되어야 한다. 그러기 위해선 학생들에게 분류할 수 있는 여러 가지 현상을 노출시켜야 하며, 분류 기능의 경험을 얻기 위해서 다양한 종류의 물체를 사용하여 분류 활동을 많이 해 보도록 격려하여야 한다(Martin, 2012, 권성기, 김동렬, 임청환 역, 2014). 여기에는 실례와 비실례가 모두 포함되어야 한다(조희형 외, 2014).

3) 교사 역할의 문제

연구 참여자들은 교사 역할의 문제를 인식하고 있었다. 이 주제는 교사 중심의 수업 반성, 지식 전달 역할의 교사 모습 반성, 교사 역할의 중요성, 과학교사의 오개념 등의 코드를 포함한다.

T14교사: 학습자 중심의 수업이 되어야 한다고 생각하면서 흉내 내기식의 모둠학습을 통해 학습자 중심의 수업으로 했다고 여기며, 여전히 교사 중심의 수업으로 이어가고 있었던 것 같다.

T17교사: 그동안 나의 과학 수업을 돌아보면 과학 자처럼 사고하고 탐구하자고 하면서 여전히 지식전달 위주의 교수학습을 하고 있었다는 반성을 하게 되었다. 학력부담이 비교적 적은 초등에서도 이렇게 지식전달이라는 교사로서의 과제에 얽매어 학생들의 탐구과정을 무시한 채 수업을 해왔었는데, 중·고등학교 선생님들께서는 훨씬 더 고민이 많으실 것이라 생각된다.

T14교사와 T17교사 모두 학습자 중심의 탐구 수업이 이루어져야 하나 여전히 교사 중심의 수업을 해왔음을 반성하고 있다. 이러한 측면에서 교육 현장의 교사 중심 수업은 개선되어야 할 것이다.

T5교사: 어쩌면 교사인 나조차도 학생들에게 가르치는 그 개념에 대해 한정적으로 알고 있

어 학생들과 진행한 탐구 활동에 대해 심도 있는 논의와 결론이 없이 끝내고 결국은 지식을 전달하는 것으로 마무리한 것은 아닐까?

T12교사: 직관적 경험에 의해 또는 학습에 의해 다년 간 배워오고 가르쳐온 과학 교사의 개념이 어떻게 구현되고 얼마나 오류투성이었는지를...

T5교사와 T12교사는 교사가 가진 한정적인 개념, 오개념에 대해 말하고 있다. 이는 교사들이 머신러닝 기반의 물질의 상태 분류 모델을 구축하는 과정에서 드러난 자신들의 개념을 인식하였기 때문이다.

나. 머신러닝 기반 교육 프로그램에 대한 교사의 인식

1) 과학 개념 생성 과정의 체험

연구 참여자들은 인공지능 기반 물질의 상태 분류 교육 프로그램에 참여하면서 과학 개념의 생성 과정을 직접 체험할 수 있었다. ‘과학 개념 생성 과정의 체험’이라는 주제는 과학적 개념의 정교화, 과학적 사고, 과학 지식의 생성 과정 체험, 관찰과 분류의 중요성, 다양한 관찰 관점과 사례·분류 기준의 중요성, 인공지능은 탐구활동의 수단, 인공지능을 통한 귀납적 탐구 경험, 합리적 추론, 과학의 본성 등의 코드어를 포함한다.

T1교사: 컴퓨팅사고와 인공지능 프로그램을 과학 교육에 어떻게 적용할지 막연했지만 직접 머신러닝, 오렌지 프로그램을 실행시켜보고 과학 개념을 귀납적으로 도출하고 합의하는 과정을 거치다 보니...

T2교사: 과학 지식(고체, 액체, 기체의 특징을 알고, 물질을 고체, 액체, 기체로 분류하는 것)보다 더 중요한 것은 체계적인 지식을 세우는 방법인 과학적 방법을 아는 것!

T3교사: 과학적 사고는 이미 답이 정해져 있는 것이 아니라 현상을 있는 그대로 관찰하고 관찰된 사실을 바탕으로 공통점과 차이점

을 찾아 현상을 분류하는 과정에서 길러진다고 생각합니다. 하지만 주어진 기준 안에서 물질의 상태를 제대로 구분하지 못하는 컴퓨터 프로그램을 통해 과학적 사고가 제대로 이루어지지 않는다는 것을 알 수 있었습니다. 정해져 있는 답안에서 짜 맞추기 하는 것이 아니라 새로운 현상이 나타났을 때 기준을 언제든지 수정, 보완될 수 있도록 수업이 이루어져야 합니다.

T1교사는 물질의 상태 분류에 대한 머신러닝 기반 학습에 임하면서 귀납적 방법으로 과학 개념을 도출하는 과정을 경험해 보았다고 말하고 있다. 머신러닝 기반 교육 프로그램은 속성과 관계, 유사성과 상이성을 파악하는 기회를 제공해주어(Klauer & Phye, 1994) 학습자들이 귀납적 사고과정을 경험할 수 있게 해준다. 그리고 T2교사는 머신러닝 기반 교육 프로그램에 참여하면서 과학 지식을 아는 것보다 학생 스스로 지식을 구성할 수 있도록 하는 방법을 아는 것이 중요함을 인식하게 되었다. 귀납적 추론과 같은 사고과정을 많은 예비 및 현직 교사들이 탐구로서 배우지 못하였고, 지식 생성 과정으로서의 탐구에 기반한 수업 경험도 부족하다는 지적이 있다(심재호 외, 2010). T2교사는 과학적 지식의 중요한 구성요소로서 과학 개념이 획득되는 과정과 방법을 교육해야 한다는 점을(권용주 외, 2013) 인식하였다. T3교사 역시 현상을 관찰하고 이를 토대로 공통점과 차이점을 찾아 분류하는 과학적 사고과정이 중요함을 언급하였다.

T1교사: 다양한 실제 사례들을 제시하여 추가로 어떤 기준들이 적합할지 생각해보고 토론하게 한다면 학교급 간의 괴리를 줄이고, 사고의 확장을 가능하게 할 것이다.

T4교사: 연역적 사고를 토대로 오류 없이 가르치는 것을 넘어서서 애매모호한 다양한 사례에 노출시켜 추론과 합의를 통해 합리적 판단을 이끌어 내도록 하는 것이 진정

한 성장이 일어날 수 있는 교육이 될 것이다.

T5교사: 그림에도 불구하고 학생들은 슬라임은 액체예요 고체예요? 하는 질문을 한다. 따라서 판단하기 애매한 기준에 대해 함께 토의하고 합의하며 답을 찾아보는 과정이 초등학생에게도 필요할 것이다.

T1교사와 T4교사 모두 물질의 상태 분류 수업에서 다양한 사례 제시를 통해 학생들이 과학적 사고가 확장되는 경험을 할 수 있을 것으로 판단하고 있다. 과학적 사고가 과학적 방법과 과학적 탐구과정을 따르는 비판적 사고의 한 유형이라는 점에서(조희형 외, 2014) 분류가 어려운 물질에 대해 느끼는 혼란을 분류 기준의 조정, 과학적 추론, 토론과 합의라는 방법을 통해 개선하고자 하는 노력은 비판적 사고에 기반한 것이며, 이를 통해 과학적 사고의 경험을 제공한다고 볼 수 있다.

T5교사의 경우는 슬라임이 액체인지 고체인지를 묻는 학생의 질문에 대한 자신의 경험을 이야기하며, 이를 수업에서 다룰 필요성에 대해 말하고 있다. 혼합물의 상태를 분류하는 것이 타당한가에 대한 논쟁이 있을 수 있으나 학생들이 물질에 대한 선개념을 형성할 때 접하게 되는 일상생활의 많은 물질들은 혼합물이다. 따라서 교사는 이러한 학생들의 의문에 대해 함께 고민해보는 시간을 가질 필요가 있다. 또한 혼합물의 분류는 관찰 사실들이 단지 수적으로 많아서는 안 되고 다양한 조건에서 얻은 사실들이어야 한다는 귀납의 원리를 고려하였을 때(권재술 외, 2013) 다양한 사례 제시는 분류 기준의 확장을 이끌어 개념을 정교화할 수 있으므로 관련 단원의 교수 전략 수립이나 교과서 등 자료 제작에 있어서 고려하여야 할 사항일 것이다.

2) 인지 갈등

연구 참여자들은 물질의 상태 분류에 대한 머신러닝 기반 교육 프로그램을 통해 인지 갈등을 느꼈으며, 머신러닝 프로그램이 인지 갈등을 일으키는 효과적인 수

단이 될 것으로 판단하고 있었다. 인지 갈등이라는 주제는 개념과 실제의 불일치, 오개념, 인지 갈등, 내가 가진 개념을 드러내는 도구, 사고의 충돌 등의 코드를 포함한다.

T4교사: 분류기준을 적용하는데 있어 일관성 없이 분류기준을 적용하는 경우가 나타났다. 그 원인을 살펴보면 분류기준을 과학자들은 미시적 세계의 순물질에서 한 종류의 입자배열로 이해하지만 우리가 일상생활에서 사용하는 개념은 혼합물이나 거시적인 세계로서 보이는 현상으로 이해하고 판단하여 분류기준을 세움으로써 이 둘 사이에서 갈등, 충돌을 가져오기 때문이다.

T7교사: 인공지능을 활용한 수업이 본인의 사고를 돌아보고 다른 사람들의 생각과 비교하며 사고를 확장해 나가는 데에 아주 효과적인 수단이 될 수 있었던 것 같다. 올해는 학생들이 깊이 궁구할 수 있는 시간을 마련해 보는 시도를 꼭 해보아야겠다.

T4교사는 일관성 없는 분류 기준 적용의 문제를 언급하면서 그 원인으로 물질을 바라보는 과학자의 미시적 세계와 거시적 현상 사이의 충돌을 들고 있다. 이것은 학습자가 학습 상황에서 겪을 수 있는 인지갈등을 의미한다. T4교사는 개인 내 인지갈등의 측면을 말하고 있다면, T7교사는 동료 간 인지갈등의 측면을 이야기 하고 있다. 즉, T7교사의 경우 머신러닝을 활용한 수업을 통해 다른 사람들과의 생각 차이를 효과적으로 비교할 수 있다는 측면에 주목하고 있다. 본 교육 프로그램에서 구축된 머신러닝 의사결정트리 모델은 나의 데이터뿐만 아니라 다른 연구 참여자들의 데이터들을 모두 귀납적으로 학습하여 생성된 것이므로 모델의 예측 결과와 나의 예측 결과의 비교를 통해 동료들의 생각을 살펴볼 수 있다. 개념 변화를 위해서는 현재 개념에 불만족을 느껴야 한다는 초기 조건을(Hewson & Hewson, 1992) 고려해보면 머신러닝 프로그램은 모델이 예측한 결과를 즉시적으로 출력하기 때문에 학습자

가 가진 생각과 바로 비교하여 봄으로써 현재 개념에 불만족을 느끼는 인지갈등을 일으킬 수 있는 수단이 될 수 있다.

T4교사: 이런 시점에서 이번 수업을 통해 ML for kids, 오렌지 프로그램을 이용하여 과학 개념을 도출하고 합의해 나가는 과정을 거치면서 인간의 뇌에서 일어나고 있는 사고 과정을 들여다 볼 수 있었고, 학습자가 어디서부터 오개념을 형성하게 되는지에 대해 알게 되었다.

T6교사: 학생들의 솔직한 인지구조를 파악할 수 있고 학생들이 개념 형성에서 도움이 되는 활동을 제공할 수 있을 것 같다.

T4교사와 T6교사 모두 물질의 상태 분류 학습에서 머신러닝 프로그램을 활용함으로써 학생들의 사고 과정을 들여다볼 수 있다는 점을 공통적으로 인식하고 있었다. 분류 활동 시 오개념이 발생하는 분기점을 파악하는데 도움이 되며, 이를 통해 과학적 개념 형성에 도움이 되는 교육 활동을 제공할 수 있으리라는 기대를 나타냈다.

3) 의사소통

연구 참여자들의 진술에서 개념 합의의 어려움, 용어에 대한 해석의 주관성, 인공지능은 의사소통의 수단, 토의·토론과 합의, 의사소통의 중요성 등의 코드를 찾을 수 있었으며, 이를 의사소통이라는 주제로 군집화 하였다.

T8교사: 과학시간에 사용되는 용어에서 항상 누구나 받아들일 수 있는 객관적인 용어를 사용하여 기준으로 정하거나 현상을 서술해야 한다고 강조하고 있지만 실제로 교사들조차 교과서에서 사용하고 있는 용어들에 대해 서로 다른 해석을 하고 있다. 생활에서 자주 사용되는 용어일수록 더욱 그러하다. 수업 시간에 무심코 사용하는

용어들에 대해 교사와 학생 또는 학생들
간에 합의의 시간이 있어야 한다.

T9교사: 머신러닝을 실행해보면서 귀납적 방법으
로 과학 개념을 정리하는 과정을 직접 경
험해보니 과학 교육에 있어 학생들의 토
론(논의)의 과정이 필수적이라는 생각이
들었다. 학생들의 토론 수준이 비전문적
이라고 하더라도 학생들 간의 토론 과정
이 선행된 후 개념 정립이 이루어져야지
교과서 개념 즉, 이미 정해져 있는 정답부
터 제시하고 그것에 현상을 끼워 맞추는
것은 교육적으로 바람직하지 않다는 생
각도 들었다.

T8교사는 물질의 상태 분류 활동을 통해 용어와 연
구 참여자들 간에 개념 사용에 있어 혼란스러움을 지
적하며, 교사와 학생 간 의사소통의 중요성을 느끼고
있었다. T9교사는 학생들의 개념 획득이 일방적 지식
전달이 아닌 탐구라는 과학 지식의 생성 과정 안에서
이루어져야 함을 말하며 이를 위해 토론의 필요성을
인식하였다. 이와 관련하여 Kuhn과 동료들은(Kuhn,
Shaw & Felton, 1997) 형식교육이 시작되는 시점부터
과학적 사고를 하는 방법, 즉 근거를 토대로 지식 주장
을 하는 사고방식을 수업에서 다루어야 한다고 주장하
였다. 그리고 이선경(2012)은 개념을 밖으로 표출하여
개념의 위상을 드러내고 그 개념이 얼마나 설득적인지
를 따져보는 과정은 교사와 학생 간, 그리고 학생들 간
의 토론 과정에서 보다 효과적임을 말하였다.

T10교사: 학생들의 과학적 탐구를 이끌어 줄 수 있
는 방법으로 인공지능 활용이 앞선 문제
의 해결방법으로 활용될 수 있을 것이라
는 생각이 들었다. 인공지능으로 나온 결
과나 과정에서 질문을 던지고 학생들이
서로 의견을 나누고 상호작용하게 하고
사고를 유발하는 좋은 도구인 것 같다.

T11교사: AI 프로그램을 통해 물질의 분류를 실천
해보고 합의한 내용과 다른 부분을 통해

재합의 또는 잘못을 범하고 있는 사고를
정정할 수 있고, AI 프로그램의 결과를
수용하는 것이 아닌 비판적 사고를 기를
수 있다.

T10교사와 T11교사는 머신러닝 모델 구축의 과정
및 머신러닝 모델이 예측한 결과를 가지고 의사소통의
소재 및 상호작용을 유발하거나 평가하는 도구로 삼을
수 있음을 말하고 있다. 토의는 상대방의 생각을 평가,
기각, 선택할 수 있기 때문에 참여자의 역할에 의해 생
각이 변화하는 것이 특징이다(Bentley & Watts, 1989).
상대의 생각을 평가하여 선택 또는 기각을 할 때 전통
적인 토의·토론 수업에서 더 나아가 머신러닝을 이용
한 과학학습 토론의 경우 입·출력 데이터, 모델의 예
측 값이 있기 때문에 상대의 논거를 평가하는데 효과
적이라 볼 수 있다.

모형-형성 문제 해결 연구 전통에서는 과학을 근본
적으로 문제 해결 활동으로 간주한다(Stewart & Hafner,
1991). 과학을 행하는 과정은 공동체 내에서 개념을 창
출하고 공유된 의미를 발전시키는 것이다. 이 관점에서
과학 학습은 설득력과 예언력을 가진 모형을 얻기 위
한 수정을 통하여 협동적으로 과학 지식을 얻는 과정
이다(Stewart & Hafner, 1991; Hafner & Stewart, 1995).
학습자가 가진 개념의 내적 일관성으로 설명할 수 없
는 현상들에 맞닥뜨리고 의사소통을 바탕으로 학습자
공동체가 인정하는 과학 지식을 생성하는 것은 쿤의
정상과학이 발전하는 양상에 비유할 수 있다. 쿤의 정
상과학이 균열이 생기는 과정은 머신러닝 모델의 예측
이 입력된 데이터와의 불일치하는 정도가 커짐과 비견
되며, 따라서 머신러닝 기반의 과학교수학습은 모델의
성능을 통해 학습자가 불만족을 느끼고 개선해나가는
조작과 동료와의 의사소통을 조율하는 효과적인 도구
가 될 수 있다.

4) 새로운 교수학습 방안

연구 참여자들은 머신러닝을 새로운 교수학습 방안
으로 인식하고 있었다. 새로운 교수학습 방안이라는 주

제는 머신러닝 모델과 인간의 예측 과정의 유사성, 새로운 교수학습 도구, 알고리즘 시각화의 유용성, 학습자 중심의 수업, 자기장학도구로써의 인공지능 등의 코드를 포함한다.

T12교사: 수업 전에는 알고리즘의 판단은 저의 머릿속 판단 과정과 다르다고 생각했습니다. 그런데 알고리즘을 통해 의사결정하는 머신러닝의 판단과, 우리의 판단을 비교해 보면 사실은 같다는 것을 알게 되었습니다. 머릿속에서 두루뭉술하게 오랜 학습과 경험이라는 애매한 함수에 의해 결정되었던 우리의 판단도, 사실은 수만 번의 물질에 대한 경험 회로를 여러 차례 반복 시행한 결과 만들어진 코드에 의해 노드를 만날 때마다 한쪽을 향해 의사 결정하는 머신러닝과 같다는 것을 확실히 이해할 수 있었습니다.

T12교사는 머신러닝 알고리즘에 대한 체험을 통해 인간이 판단을 내리는 과정과의 유사성을 깨닫고 있었다. 이는 컴퓨팅 사고라 불리는 것으로 SW교육 등에서 강조되어 왔으나 과학교사로서 쉽게 체감하지 못하였던 것을 머신러닝 기반 교육 프로그램을 통해 이해하게 된 것으로 볼 수 있다.

T6교사: 과학적 개념을 이해하는 데 필요한 속성을 고민하고 0과 1로 분류하는 것만으로도 학습이라고 생각한다. 왜냐하면 평소에는 지나쳤던 것들인데 고민했던 속성에 대해서는 한 번 더 주의를 기울이게 되었다.

T10교사: 프로그램을 이용하여 시각적으로 체계화할 수 있어 개념들을 도식화하는데 도움이 될 것 같다는 생각이 들었다.

T6교사는 머신러닝 알고리즘이 학습할 데이터를 생성하는 과정에서 분류 기준에 따라 0, 1로 속성 값을

부여하는 것으로도 충분한 학습이 되었다고 말하고 있다. 분류라는 탐구가 수행되는 과학 학습에서 머신러닝이 새로운 교수학습 도구로 사용될만한 가치가 있다고 평가한 것이다. 기존의 물질의 상태 분류 학습은 세부적인 분류 기준을 어떻게 적용하였는지에 대한 학습자의 생각을 명시적으로 드러내지 않고, 물질의 정의에 입각한 총체적인 판단으로 수행되는 경향이 있기 때문이다. 그리고 T10교사는 머신러닝 알고리즘 시각화 기능이 학습자의 개념을 도식적으로 드러낼 수 있다는 점을 언급하고 있다. 본 연구에 한정짓는다면 의사결정트리의 시각화를 통해 분류의 과정을 살펴보고, 어느 분기점에서 나와 생각이 달랐는지를 살펴볼 수 있기 때문에 과학교과의 분류 내용 교수학습에 유용할 것이다.

5) 시너지 효과

시너지 효과라는 주제는 과학교육과 머신러닝의 접목을 통해 얻을 수 있는 긍정적인 효과를 의미한다. 여기에는 메타인지, 미래 역량, 발산적 사고 촉진 등이 포함된다.

T6교사: 인지구조를 드러내는 도구로써 무엇을 알고 모르는지 알 수 있어 학생들의 메타인지를 활용하여 자기관리 역량을 키울 수 있을 것 같다.

T13교사: 머신러닝 AI를 통한 수업을 통해 앞으로 필요한 교수방법을 선도적으로 배워 학생들에게 전해야겠다는 마음이 들었고, 하나를 알더라도 깊이 있고 발산적 사고를 촉진할 수 있도록 노력해야겠습니다. 충분한 의사소통과 같이 학생들에게 개념이 형성되기까지 많은 담론이 필요하며 이를 통해 더욱 구체화하고 명료화한 개념을 비판적으로 학습 가능하다는 것을 알게 되었습니다. 앞으로의 수업에는 발산적 질문 및 학생과 소통하는 수업으로 진행할 필요를 느꼈습니다.

과학적 사고의 발달에 중요한 메타인지는 내가 믿고

있는 것과 나에게 제시되고 있는 증거를 구별하고, 그 둘 간의 관련성을 파악하여 증거가 얼마나 믿을만한지를 의식적으로 통제하는 것이다(이선경, 2012). 머신러닝 모델을 구축하기 위한 과정 및 모델의 예측 결과의 비교 및 조정은 나의 생각과 모델의 판단을 명시적으로 드러냄으로써 메타인지 활동을 수반하게 된다고 볼 수 있다. 이러한 점에서 T6교사는 머신러닝 기반의 교수학습을 통해 학습자 스스로 무엇을 알고 무엇을 모르는지를 인식하게 함으로써 메타인지가 활용될 수 있음을 인식하였다.

T13 교사의 경우 물질의 상태 분류에 대한 머신러닝 기반 교육 프로그램 참여를 통해 발산적 사고가 촉진되는 과학 수업의 가치를 인식하였다. 즉, 명확한 상태 분류가 어렵지만 학생들이 일상에서 접하게 되는 대부분의 물질인 혼합물을 분류하려는 교육적 시도 및 분류를 위한 다양한 기준의 생성, 그리고 그것들을 비판적으로 다루는 과학적 담화의 필요성을 인식하였다.

6) 머신러닝 적용의 걸림돌

연구 참여자들은 머신러닝을 과학교육에 적용하는 것에 대한 긍정적인 인식과 더불어 현장 도입의 어려운 점에 대해서도 함께 언급하였다. 이를 머신러닝 적용의 걸림돌이라는 주제로 표현하였으며, 이 주제에는 어려운 머신러닝, 머신러닝 활용교육을 위한 물리적 환경 구축 선행, 시간적 여유 필요, 머신러닝과 과학학습 내용과의 연계 방법, 교사의 머신러닝 수업 역량 필요 등의 코드가 포함된다.

T6교사: 수업을 듣기 전에는 인공지능이 막연하고 멀리 있는 것이라고 생각을 했고 잘 모르니까 멋있는 거라고 여겼다. 검은 화면에 알 수 없는 규칙에 따라 영어 문자를 작성해야 하는 것으로만 생각했고 어렵고 내가 할 수 있는 영역이 아니라고 여겼다.

T6교사는 인공지능을 다룬다는 것에 대해 복잡한 프로그래밍 언어를 배운다는 것으로 인식하여 비전공자

에게는 어려운 분야라고 여기고 있었다. 그러나 인공지능 기반 교육에 대한 인식과 이해도를 갖추면 복잡한 프로그래밍 언어를 몰라도 인공지능을 고려한 교육과정 개발이 가능하다(김현진 외, 2020). 그리고 테크놀로지 활용 학교교육에서 성공을 좌우하는 가장 중요한 요인은 교사라는 점에서(Ertmer, 1999) 인공지능에 대한 기초 소양을 기를 수 있는 교사교육이 필요하다.

T16교사: 인공지능 수업을 실제 수업에 적용하기 위해서는 충분한 시간적 여유가 있어야 한다. 또한, 동학년 교사와의 합의가 필요하고 전체 학년을 수용할 수 있는 공간과 컴퓨터가 필요하다. 현실에서 적용하기에 여러 가지 어려움이 있다면, 진도의 부담이 없는 실험 동아리를 운영하면서 적용해보면 좋을 것 같다. 무엇보다 가장 큰 고민이라면 학습한 ML for kids와 orange3를 어떤 단원에 어떻게 전달할 지이다.

T18교사: 정규 수업 과정의 한계 → 동아리 활동, 자유학년제 주제선택

T16교사는 인공지능을 활용한 과학수업이 이루어지기 위해서는 동료교사와의 공감대 형성 및 합의, 물리적 환경 구축, 시간 확보 등이 필요함을 말하였다. 그리고 정규 수업과정에서의 적용이 어렵다면 동아리 활동에 적용해볼 수 있음을 이야기하였으며, T18교사는 이에 더해 자유학년제 주제 선택을 통해 적용 가능성을 인식하였다. 하지만 T16교사가 말한 바와 같이 인공지능을 어떤 학습 내용에, 어떤 방식으로 적용할지에 대한 문제가 남게 된다. 이를 교사 개인의 역량에 맡기기 보다는 인공지능 기반 교사교육 프로그램을 통한 교사의 역량 강화와 함께 관련 전문가들이 교수학습 프로그램을 개발하여 보급함으로써 과학교육과 머신러닝의 접목을 확장시키는 노력이 필요하다.

IV. 결론 및 제언

이 연구에서는 초·중등 과학교사 대상의 물질의 상태 분류에 대한 머신러닝 기반 교사교육 프로그램을 개발하고 실행하여 물질의 상태 분류 머신러닝 모델을 구축하였다. 그 결과, 총 3회에 걸친 분류 활동에서 생성된 머신러닝 모델의 성능 평가를 통해 교사교육 프로그램이 교육적 효과가 있음을 확인하였으며, 질적 연구 분석을 통해 과학교사들이 머신러닝의 기반 교육 프로그램의 인식에 대한 9가지 주제를 도출할 수 있었다. 이 9가지 주제는 기존 교육에 대한 문제점 인식과 머신러닝 교육 프로그램의 인식으로 분류하였다. 이와 같은 연구 결과의 분석을 통해 내린 결론은 다음과 같다.

첫째, 머신러닝을 과학교육의 탐구학습의 방안으로 교사교육에 적극 도입할 필요가 있다. 현행 과학과 교육과정의 내용 체계에서 강조하고 있는 자료의 수집·분석 및 해석, 수학적 사고와 컴퓨터 활용, 모형의 개발과 사용의 측면에서 교사들이 수업역량을 획득하기 위해서는 머신러닝 기반 교육 프로그램이 그 대안이 될 수 있다. 머신러닝 모델은 새롭게 입력되는 데이터들에 대한 의미 있는 예측치를 제공하고, 데이터와 예측의 일치 정도를 비교, 분석 및 논증하는 평가의 과정을 가질 수 있도록 도움을 줄 수 있기 때문이다. 또한 분류를 위한 머신러닝 기반 과학학습은 학습자가 데이터를 생성하는 과정에서 분류를 위한 기준을 생성하고 대상을 관찰하여 분류 기준에 대응하는 속성 값을 입력하여 분류의 결과인 레이블을 부여하는 귀납적 탐구과정을 경험할 수 있다는 측면이 있다. 연구 참여자들인 과학 교사들의 목소리를 들었을 때 학교 현장에서의 과학수업은 여전히 지식 위주의 수업이 이루어지고 있었다. 과학의 지식과 방법은 서로 결합되어 있으며 다른 하나를 제외하고 가르치는 경우 과학의 본성에 대한 학습자의 이해를 왜곡시키게 된다(Chiappetta & Koballa, 2006). 따라서 인공지능 시대의 미래 인재를 기르기 위한 시대적 요구를 고려할 때 머신러닝 기반 교육 프로그램의 도입은 교사뿐만 아니라 학생에게도 필요하다.

따라서 앞으로 학생들을 위한 머신러닝 교육 프로그램의 개발과 그 교육적 효과를 알아보는 연구도 수행되어야 할 것이다.

둘째, 과학교사들의 물질의 상태 분류 활동은 지식에 근거한 연역적 사고과정이 영향을 미치는 것으로 나타났다. 이러한 연역적 사고과정은 머신러닝 의사결정트리 모델의 분류 결과와의 차이를 야기하였다. 구성주의적 과학 지식을 생성하기 위해서 논증, 비평, 분석을 통한 평가의 과정을 거쳐야 하는데 연역적 사고를 기반으로 정답을 부여하는 식으로 분류활동에 대한 교육이 진행된다면 논증, 비평, 분석의 기회는 사라질 것이다. 따라서 귀납적 사고과정과 연역적 사고과정을 질적으로 다른 동등한 사고활동으로 이해하고, 서론에서 살펴본 과학자와 공학자의 활동 범주 프레임 워크에서 귀납적 사고에 근거한 데이터 수집은 학습자가, 그리고 알고리즘이 데이터를 귀납적으로 학습하여 구축한 모델을 통해 예측과 설명을 이끌어내는 연역 추론을 수행하고, 데이터와 예측 결과의 비교를 통해 논증, 비평, 분석이 이루어질 수 있도록 하는 과정을 교사가 이해하고 과학교육에 적용할 필요가 있다.

셋째, 미시적 분류 기준이 물질의 상태 분류에 있어서 거시적인 분류 기준보다 상대적으로 영향을 주지 않았다. 가장 정교화 된 Ver. 3 모델의 성능을 미시적 분류 기준을 포함하였을 때와 제외하였을 때를 비교한 결과 F1-Score의 값이 각각 0.908, 0.911로 나타났다. 또한 어떤 분류 기준이 규칙 트리를 만드는데 기여하였는지를 나타내는 특성 중요도 값 역시 거시적 분류 기준들이 미시적 분류 기준에 비해 상대적으로 더 높은 특성 중요도 값을 나타냈다. 이것은 분류라는 탐구 활동이 관찰에 의해 파악할 수 있는 거시적 특성의 속성으로 귀납적 추론을 통해 무리를 짓는 측면을 가지고 있기 때문이다. 분자의 배열 상태와 같은 미시적 분류 기준의 경우 관찰이 불가능한 연역적 속성이다. 따라서 관찰 가능한 속성인 거시적 분류 기준만으로 물질의 세 가지 상태에 대한 분류 결과가 정교화 되었다는 본 연구의 결과는 미시적 분류 기준이 절대적으로 중요하게 다루어지는 중학교 이상의 과학교육 내용 구

성에 주는 시사점이 크므로 후속연구를 통해 더욱 면밀히 살펴볼 필요가 있다.

이상의 연구 결론을 바탕으로 다음과 같은 제언을 한다.

첫째, 머신러닝 기반 교육 프로그램에서 정교화된 모델을 토대로 교과서 등의 교육 자료가 개발되어야 한다. 분류는 사물의 특성을 이해하는 가장 기본적인 방법으로(Kaplan, 1964) 유사점과 준거 속성에 따라 사물과 현상을 함께 묶거나 관계를 짓는 과학적 탐구과정이다(조희형 외, 2014). 이러한 분류는 귀납적 사고과정을 수반하게 되는데 본 연구에서는 머신러닝 기반 교육 프로그램을 통해 귀납적 사고과정을 통한 분류 활동이 체계적으로 이루어질 수 있음을 확인하였다.

둘째, 혼합물을 물질의 상태 분류 학습에 포함시켜야 하는가에 대해서 과학적 관점에서의 논의가 필요하다. 개념의 전형성을 띤 분명한 사례만을 학습하는 것은 과학적 추론과 분석, 논증의 과정 등을 통해 지식을 생성하는 과학 활동을 학생들이 경험하는 것에 도움이 되지 못한다. 물질은 질량을 가지고 공간을 차지하는 것으로 고체, 액체, 기체의 세 가지 상으로 존재한다고 설명된다(Masterton & Hurley, 2004, 일반화학교재연구회 역, 2005). 그러나 세상의 많은 물질들이 순물질이 아닌 상태로 존재하고 있으며, 다양한 물질을 세 가지 상태만으로 상태 분류하는 것은 어려운 일이다. 본 연구에서도 응답 집중 지수를 사용하여 물질의 28가지 사례 별 분류 결과를 살펴본 결과 응답의 분산 정도가 큰 사례들은 대부분 혼합물이었다. 예를 들어 공기 중에 액체 분산질이 부유하고 있는 구름의 경우 그것을 구성하고 있는 성분에 초점을 두느냐 또는 성분이 포함된 계에 초점을 둔 혼합물 상태를 바라보느냐에 따라 상태 분류 결과가 달라졌다. 교사의 경우에서도 이러한 혼란이 나타나므로 학생들의 경우에는 이러한 문제가 더욱 심각할 것이다. 이는 선행연구에서(최정인, 백성혜, 2015) 이미 확인되었다. 그러나 학생들의 호기심의 대상은 분류가 명확한 순물질에 국한되지 않으며, 다양한 혼합물도 포함될 것이다. 따라서 일상생활에서 체험할 수 있는 혼합물이 물질의 상태에 대한 비전형

성을 띤다고 하더라도 과학적 추론과 분석, 논증의 과정 등을 통해 지식을 생성하는 과학 활동을 학생들에게 경험시키기 위하여 머신러닝 기반 교육 프로그램을 활용하여 분류 활동에 포함시키는 것을 제안한다.

참고문헌

- 가석현, 이민구, 김찬중 (2020). 지능정보사회를 위한 피지컬컴퓨팅 융합 과학교육 제안. **한국컴퓨터교육학회 학술발표대회논문집**, 24(1), 83-87.
- 강경희 (2013). **과학교육 프로그램 개발의 이론과 실제**. 파주: 한국학술정보(주).
- 강태정 (2000). 물질의 상태와 상태변화에 대한 초등학생의 개념 조사. 한국교원대학교 교육대학원 석사학위논문.
- 교육부 (2015). **과학과 교육과정**. 서울: 교육부.
- 권용주, 남정희, 이기영, 이효령, 최경희 (2013). **과학교육 사고에서 학습까지**. 서울: 북스힐.
- 권용주, 정진수, 신동훈, 이준기, 이일선, 변정호 (2011). **과학적 탐구력 향상을 위한 과학지식의 생성과 평가**. 서울: 학지사.
- 권재술, 김범기, 강남화, 최병순, 김효남, 백성혜, ... 정진우 (2013). **과학교육론**. 파주: 교육과학사.
- 권철민 (2020). **파이썬 머신러닝 완벽 가이드**. 파주: 위키북스.
- 김선경, 김영미, 백성혜 (2008). 다양한 물질의 상태에 대한 중고등학생들과 과학교사들의 분류 기준에 대한 분석. **대한화학회지**, 52(5), 569-579.
- 김영환 (2021). 인공지능교육 관련 연구주제 분석. **교육혁신연구**, 31, 197-217.
- 김진규, 윤길근, 박형근 (2021). **인공지능 시대의 인재 혁명**. 파주: 교육과학사.
- 김현진, 박정호, 홍선주, 박연정, 김은영, 최정윤, 김유리 (2020). 학교교육에서 AI 활용에 대한 교사의 인식. **교육공학연구**, 36(3), 905-930.
- 류혜인, 고아라, 조정원 (2019). 초중등에서의 인공지능 교육 방향 제언. **한국정보과학회 학술발표논문집**, 763-765.
- 박민술, 박주본, 박유민, 조영주 (2020). 미래 지능형 과학실 활용을 위한 “화학원소기호 이미지 기계학습 AI·SW 교육 프로그램” 제안. **한국컴퓨터정보학회 학술발표논문집**, 28(2), 629-632.
- 박성익, 임철일, 이재경, 최정임, 임정훈, 정현미, ... 이지은 (2012). **교육공학의 원리와 적용**. 파주: 교육과학사.
- 백성혜, 최정인, 박은주 (2013). 초등학교 3학년 ‘물질의 상태’ 단원에 제시된 예의 기능별 유형 분석 및 학생들의 이해. **한국과학교육학회지**, 33(7), 1273-1284.
- 성현란 (1989). 학령 전 아동에서의 자연범주와 귀납적 추론. **한국심리학회지**, 2(1), 18-31.

- 송진웅, 강석진, 박영순, 김동건, 김수환, 나지연, ... 김진희 (2019). **모든 한국인을 위한 과학적 소양 미래세대 과학교육표준**. 한국과학창의재단.
- 신원섭 (2020). 운동과 에너지에서 인공지능 융합 과학교육의 가능성 탐색 연구. **에너지기후변화교육**, 10(1), 73-86.
- 신원섭, 신동훈 (2021). 온라인 학습에서 머신러닝을 활용한 초등 4 학년 식물 분류 학습의 적용 사례 연구. **초등과학교육**, 40(1), 66-80.
- 심재호, 신명경, 이선경 (2010). 2007 개정 과학과 교육과정의 주요 내용의 실행에 관한 현직 과학 교사의 인식. **한국과학교육학회지**, 30(1), 140-156.
- 유주경 (2012). 초등학교 과학 교과서에 제시된 개념에 대한 예의 기능적 유형 분석. **청주교육대학교 교육대학원 석사학위논문**.
- 이경진, 하희수, 홍훈기, 김희백 (2018). 머신 러닝을 활용한 과학 논변 구성 요소 코딩 자동화 가능성 탐색 연구. **한국과학교육학회지**, 38(2), 219-234.
- 이선경 (2012). **과학 교사의 수업 전문성: 연구와 실행**. 파주: 교육과학사.
- 이선경 (2015). **과학학습 개념변화**. 서울: 서울대학교출판문화원.
- 이소율, 이영준 (2021). 머신러닝 교육 플랫폼 활용 ‘분자 구조의 이해’를 위한 융합교육 프로그램 개발. **정보교육학회논문지**, 25(6), 961-972.
- 장병철, 전수진, 김한성, 한창수, 이현아, 김학인, ... 장의덕 (2021). 고등학교 학생을 위한 인공지능 교육자료 개발. **한국컴퓨터교육학회 학술발표대회논문집**, 25(1(A)), 37-40.
- 정영식 (2015). 초등학교를 위한 알고리즘 및 프로그래밍 교육과정 모델 개발. **한국정보교육학회지**, 19(4), 459-466.
- 조희형, 김희경, 윤희숙, 이기영 (2014). **과학 교재연구 및 지도법**. 파주: 교육과학사.
- 최경숙 (2003). 물질의 상태와 상태변화에 대한 중학교 1학년 학생들의 개념조사. **한국교원대학교 교육대학원 석사학위논문**.
- 최연구 (2017). 4차 산업혁명시대의 미래교육 예측과 전망. **Future Horizon**, 33, 32-35.
- 최정인, 백성혜 (2015). 물질의 세 가지 상태에 대한 개념 변화에 초등교사들의 토론이 미치는 영향과 분류활동 결과의 분석. **대한화학회지**, 59(4), 320-335.
- 최정인, 백성혜 (2016). 초등학교 3-6학년 학생들의 물질의 상태 개념 속성의 특징. **대한화학회지**, 60(6), 415-427.
- 한지윤, 신영준 (2020). 인공지능교육 관련 연구 동향 분석: 키워드 네트워크 분석. **인공지능연구 논문지**, 1(2), 20-33.
- 홍선주, 조보경, 최인선, 박경진 (2020). **학교교육에서 인공지능(AI)의 개념 및 활용** (연구자료 ORM 2020-21-3). 한국교육과정평가원.
- Bao, L., & Redish, E. F. (2001). Concentration analysis: A quantitative assessment of student states. *American Journal of Physics*, 69(S1), S45-S53.
- Bentley, D., & Watts, M. (1989). *Learning and teaching in school science: Practical alternatives*. Maidenhead: Open University Press.
- Chiappetta, E. L., & Koballa, T. R. (2006). *Science instruction in the middle and secondary schools: Developing fundamental knowledge and skills for teaching*. NJ: Pearson Prentice-Hall.
- Council, N. R. A. (2012). *Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. Washington, D.C.: National Academies Press.
- Dangeti, P. (2017). *Statistics for Machine Learning*. Birmingham: Packt Publishing Ltd.
- Ertmer, P. A. (1999). Addressing first-and second-order barriers to change: Strategies for technology integration. *Educational Technology Research and Development*, 47(4), 47-61.
- Hafner, R., & Stewart, J. (1995). Revising explanatory models to accommodate anomalous genetic phenomena: Problem solving in the “context of discovery”. *Science Education*, 79(2), 111-146.
- Hewson, P. W., & Hewson, M. G. (1992). The status of students’ conceptions. *Research in physics learning: Theoretical issues and empirical studies*, 59-73.
- Hurwitz, J., & Kirsch, D. (2018). *Machine learning for dummies: IBM Limited Edition*. NY: John Wiley & Son Inc..
- Kaplan, A. (1964). *The Conduct of Inquiry: Methodology for Behavioral Science*. San Francisco: Chandler Pub. Co.
- Klauer, K. J., & Phye, G. D. (1994). *Cognitive training for children: A developmental program of inductive reasoning and problem solving*. Seattle: Hogrefe & Huber.
- Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentative reasoning. *Cognition and Instruction*, 15(3), 287-315.
- Madison, D. S. (2005). *Critical ethnography: Methods, ethics, and performance*. Thousand Oaks, CA: Sage.
- Martin, D. J. (2012). *Elementary science methods: A constructivist approach*. Cengage Learning.
- 권성기, 김동렬, 임청환 역 (2014). **초등교사를 위한 과학교육**. 서울: 북스힐.
- Masterton, W. L., Hurley, C. N. (2004). *Chemistry: Principles and Reactions* (5th ed.). 일반화학교재연구회 역 (2005). **일반화학**. 파주: 자유아카데미.
- Matthews, M. R. (1994). *Science teaching: The role of history and philosophy of science*. 권성기, 송진웅, 박종원 역 (2014). **과학교육: 과학사와 과학철학의 역할**. 파주: 북스힐.
- Müller, A. C., & Guido, S. (2016). *Introduction to machine*

- learning with Python: A guide for data scientists*. Sebastopol: O'Reilly Media Inc.
- Orange (2021, January 25). *Orange*. Retrieved January 25, 2021, from <https://orangedatamining.com/>
- Rampin, R., & Rampin, V. (2021). Taguette: Open-source qualitative data analysis. *Journal of Open Source Software*, 6(68), 3522.
- Richmond, G., & Striley, J. (1996). Making Meaning in Classrooms: Social Processes in Small group Discourse and Scientific Knowledge Building. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 33(8), 839-858.
- Stepans, J. I. (2003). *Targeting students' science misconceptions: Physical science concepts using the conceptual change model*. Tampa: Showboard.
- Stewart, J., & Hafner, R. (1991). Extending the Conception of “Problem” in Problem-Solving Research. *Science Education*, 75(1), 105-20.
- Taguette (2022, August 26). *Taguette*. Retrieved August 26, 2022, from <https://www.taguette.org/>
- World Economic Forum (2016). *The Future of Jobs: Employment, Skills and Workforce Strategy for the Fourth Industrial Revolution*. Geneva: The Author.

ABSTRACT

The Educational Effect of Machine Learning-Based Teacher Education Program on the Classifying State of Matter

Jung-In Choi¹, Seoung-Hey Paik²

¹Teacher, Incheon Dangsang Elementary school,

²Professor, Dept. of Chemical Education, Korea National University of Education

Objectives In this study, a machine learning-based teacher education program for classifying the state of matter was developed and implemented for elementary and secondary science teachers and machine learning model for state classification of matter was established. And by analyzing the model evaluation and the experience of science teachers qualitatively, it was attempted to confirm the effectiveness of the program and the perception of science teachers about the application of machine learning to science classes.

Methods For 31 elementary and middle school science teachers enrolled in the Graduate School of Education at the College of Education located in the central region, a total of three matter classification activities were performed and a decision tree algorithm was applied to the machine learning model. And the effectiveness of the program was confirmed through model performance evaluation such as accuracy and F1-score. In addition, we qualitatively analyzed the thoughts of science teachers on the application of machine learning.

Results As a result of evaluating the performance of the built model during a total of 3 state of matter classification activities, it was determined that the accuracy and F1-score values increased statically, resulting in the educational effect of the Machine Learning-Based Teacher Education Program on the Classifying State of Matter. And as a result of analyzing the data of the study participants, it was revealed that there are cases in which the states of matter are classified through a deductive thinking process based on knowledge. As a result of comparing evaluation indicators of the machine learning models, the macroscopic classification criteria had a greater effect than the microscopic classification criteria. And when the classification results between the machine learning model and the research participants for various examples were compared, in the case of mixtures, the degree of inconsistency increased. In addition, through qualitative research analysis, the thoughts of science teachers on the application of machine learning were derived into nine topics, including 'experience the process of creating science concepts', 'cognitive conflict', and 'communication'.

Conclusions Machine learning should be actively introduced into teacher education as an inquiry-learning method of science education. And educational materials such as textbooks that can properly reflect scientific reasoning processes such as inductive thinking should be developed based on machine learning-based education programs. And it is necessary to discuss from a scientific point of view whether mixtures should be included in the learning of Classifying State of Matter.

Key words Machine learning, Artificial intelligence, State of matter, Classification, Science teacher, Teacher education