



머신 러닝을 활용한 과학 논변 구성 요소 코딩 자동화 가능성 탐색 연구

이경건, 하희수, 홍훈기*, 김희백
서울대학교

Exploratory Research on Automating the Analysis of Scientific Argumentation Using Machine Learning

Gyeong-Geon Lee, Heesoo Ha, Hun-Gi Hong*, Heui-Baik Kim
Seoul National University

ARTICLE INFO

Article history:

Received 22 February 2018

Received in revised form

16 April 2018

23 April 2018

Accepted 23 April 2018

Keywords:

Scientific Argumentation,
Machine Learning, Artificial
Intelligence, Automation,
Natural Language Processing

ABSTRACT

In this study, we explored the possibility of automating the process of analyzing elements of scientific argument in the context of a Korean classroom. To gather training data, we collected 990 sentences from science education journals that illustrate the results of coding elements of argumentation according to Toulmin's argumentation structure framework. We extracted 483 sentences as a test data set from the transcription of students' discourse in scientific argumentation activities. The words and morphemes of each argument were analyzed using the Python 'KoNLPy' package and the 'Kkma' module for Korean Natural Language Processing. After constructing the 'argument—morpheme:class' matrix for 1,473 sentences, five machine learning techniques were applied to generate predictive models relating each sentences to the element of argument with which it corresponded. The accuracy of the predictive models was investigated by comparing them with the results of pre-coding by researchers and confirming the degree of agreement. The predictive model generated by the k-nearest neighbor algorithm (KNN) demonstrated the highest degree of agreement [54.04% ($\kappa = 0.22$)] when machine learning was performed with the consideration of morpheme of each sentence. The predictive model generated by the KNN exhibited higher agreement [55.07% ($\kappa = 0.24$)] when the coding results of the previous sentence were added to the prediction process. In addition, the results indicated importance of considering context of discourse by reflecting the codes of previous sentences to the analysis. The results have significance in that, it showed the possibility of automating the analysis of students' argumentation activities in Korean language by applying machine learning.

1. 서론

전통적인 과학 교육은 과학 개념의 학습에 치중하여 온 측면이 있었으나 실제 과학자 공동체가 지식을 구성하는 인식적 측면이 과학 교육에 포함되어야 한다는 주장이 지속적으로 제기되면서(Duschl, 2008; Lederman, 1992; MOE, 2015) 과학자 공동체의 지식 구성 과정의 핵심을 이루는 논변 활동이 도입될 필요가 강조되어 왔다(Driver, Newton, & Osborne, 2000; Jiménez-Aleixandre & Erduran, 2008). 이러한 맥락에서 과학자들이 지식 주장을 정당화하고 검증하는 과정인 논변 활동을 과학 교실의 교수학습 과정에 도입하기 위한 시도들이 있었다(e.g. Berland & Hammer, 2012; Lee, Yun, & Kim, 2015; Simon, Erduran, & Osborne, 2006). 이때 교사나 연구자들은 다양한 조건 하에서 학생들의 논변 활동을 지원하고 탐색하기 위해 학생들이 구성한 논변의 구성요소 및 구조를 분석한 바를 바탕으로 과학적 개념, 인식적 수준, 논변을 구성하는 사회적 과정 등의 여러 측면에서 분석해왔다(e.g. Erduran, Simon, & Osborne, 2004; Larrain, Freire, & Howe, 2014; Lee, Yun, & Kim, 2015).

논변 활동에서 학생들이 구성한 논변의 구성 요소를 분석하는 과정

은 학생들의 논변 활동을 탐색하는 대다수의 연구에서 기본적으로 이루어지며 논변 활동에서의 교수 전략에 있어 밑바탕이 되기에, 근래에는 학습 과정에서 수집한 정보를 토대로 교수자와 학생을 지원하는 학습 분석(learning analytics)에 머신 러닝을 도입하여(Pardo, 2014) 학생들의 담화를 분류 및 코딩하고자 하는 시도가 이루어지고 있다(e.g. McAlister *et al.*, 2014; Scheuer *et al.*, 2012; Rosé *et al.*, 2008). 예컨대 Rosé *et al.* (2008)은 학생들의 발화를 머신 러닝을 통해 코딩한 결과와 연구자의 코딩 사이의 일치도를 높이기 위한 방안을 탐색하였고, 그로 인하여 생성된 분류 알고리즘으로 발화의 인식적, 사회적 특성을 분류한 결과 다양한 머신 러닝 기법들 중 서포트 벡터 머신(Support Vector Machine, SVM)의 경우에 그 일치도가 가장 높음을 보고하였다. McAlister *et al.* (2014)은 특정 주제에 관한 방대한 양의 기존 문헌 데이터를 바탕으로 하여 특정 논변 구조를 암시하는 용어를 찾아내고, 이를 바탕으로 논변의 구조를 나타내는 프로그램을 고안하였다. 한편 Sioni *et al.* (2011)은 교실에서의 '교류적 담화(transactive discourse)'를 자동으로 분석하고 코딩하기 위한 틀을 제안하기도 하였다. 이러한 연구들은 오늘날 교육학계의 화두 중 하나인 인공지능 시대의 학습에 있어서(Heo *et al.*, 2017; Park

* 교신저자 : 홍훈기 (hghong@snu.ac.kr)

http://dx.doi.org/10.14697/jkase.2018.38.2.219

& Shin, 2017; Shin, Ha, & Lee, 2017) 머신 러닝을 활용할 수 있는 가능성을 보여준다.

기술적으로 머신 러닝은 인공지능의 하위 분야로서(Domingos, 2015), 빅 데이터 및 데이터 마이닝과의 밀접한 연관 하에서 다양한 형태의 데이터를 기계에게 학습시켜 프로그램이나 알고리즘 등의 모델을 자동으로 생성하도록 할 뿐 아니라 새로운 데이터에 관한 예측을 하는 데에 사용될 수 있다(Lantz, 2013; Lim, 2015; Zhang, 2017). 이러한 머신 러닝이 교수학습 과정에 적용되면서, 인공지능이 이미 개별 학습자를 위한 개인교사(tutor)가 되고, 학생들의 협력학습을 지원하며, 가상현실에서의 실제적인 학습을 지원하는데 사용할 수 있는 수준으로 발전하였다. 또한 앞으로는 학생들이 21세기 학습 역량을 갖도록 돕고, 형성적인 평가를 용이하게 하며, 학습과학의 발전 및 평생교육 지원에 있어서도 큰 역할을 할 것으로 기대되고 있다(Luckin *et al.*, 2016). 예컨대, 머신 러닝을 통해 학생들의 과학 논변 활동을 분석하는 기술이 이상적으로 구현된다면 교실 현장에서 즉각적으로 많은 학생들의 논변 활동을 모니터링하고 보고하여 교사가 학생들의 요구와 어려움이 무엇인지 파악하는 데에 드는 시간과 노력을 상당히 감소시켜주며 이에 적합한 지원을 제공할 수 있도록 도울 수 있을 뿐 아니라, 교사의 개입 전후에 학생들의 논변 활동 변화 양상을 파악하는 것 또한 용이해질 것이다.

하지만 위에서 살펴본 바와 같은 연구 결과들을 국내의 과학 교육 맥락에 곧바로 적용하여 이해하기에는 무리가 있다. 먼저는 머신 러닝을 통하여 담화를 자동으로 분석하는 연구들이 주로 영미권에서 이루어져왔기 때문이다. 따라서 기존 영미권 연구 결과를 우리나라 과학 교육 맥락에 반영할 경우, 기본 어순, 문법, 화용상의 원칙 등이 상이한 한국어로 이루어지는 담화(Jung, & Seo, 1998; Lee, 2002)를 분석하는 작업에서 기대와는 다른 결과를 낼 가능성이 있다. 또한 대부분의 연구가 온라인상에서 이루어진 논변 활동으로부터 데이터를 추출하였기에, 이는 음성 발화로 진행되는 논의와는 그 특성이 달라(Jang, 2003) 교실 수업에서 이루어진 논변 활동을 분석하는 데에 적용하기에는 한계를 지닌다. 국내에서 학생들의 자연어 발화를 자동으로 인식하거나 분석하려는 시도들도 있었으나, 이는 대부분 영어 등의 언어 교육 맥락에서 이루어졌으며(e.g. Ok, 2004; Lee, Lee, & Lee, 2010) 과학 교수학습 과정에서 발화된 자연어를 처리하는 연구는 보고된 바가 거의 없다. 또한 현재까지 국내에서 과학 논변 활동의 분석 방법을 고찰한 연구는 주로 논변 활동의 서로 다른 측면을 포착하기 위해서 이루어져왔으며(e.g. Lee, Maeng, & Kim, 2008; Maeng, Park, & Kim, 2013), 그 분석 및 코딩을 자동화하려는 시도는 현재까지 거의 이루어지지 않았다. 한편 머신 러닝 및 인공지능을 교육적 맥락에서 논의한 국내연구로서는 인공지능 시대의 학습에 관한 철학적 논의나(Heo *et al.*, 2017) 인공지능 또는 인공지능 교사에 관하여 초중고 학생들의 인식을 조사한 연구(Park & Shin, 2017; Shin, Ha, & Lee, 2017), 그리고 인공지능 및 머신 러닝 자체를 학습하는 방법에 관한 연구(Kim & Park, 2017; Lee, Nam, & Shin, 2017) 정도가 이루어졌으나, 데이터를 수집하고 직접 머신 러닝을 수행하여 학습에 관한 모델을 만들고 학습 분석을 수행하는 연구 역시 현재까지는 보고된 바가 거의 없다. 머신 러닝과 인공지능을 교실에 도입하는 국제적인 추세 속에서, 논변 활동을 도입하고자 하는 국내 과학 교육 맥락에 이러한 테크놀로지가 실질적인 도움을 제공

할 수 있기 위해서는 이와 관련한 연구와 논의가 선행되어야 필요가 있다.

이에 본 연구에서는 머신 러닝을 통해 우리나라의 과학 교실에서 이루어지는 논변 활동 담화에서 논변의 구성 요소를 자동으로 코딩할 수 있는 프로그램의 개발 가능성을 탐색해보고자 하였다. 먼저 이론적 배경을 고찰한 이후, 머신 러닝에 쓰일 데이터를 구축하기 위해 기존 문헌에서의 Toulmin의 틀에 따라 논변 구성 요소를 코딩한 결과를 수집하고 이에 관한 기술 통계를 확보하고자 하였다. 그리고 실제 교실 환경에서 발화된 과학 논변 활동 자료를 자연어 처리를 통해 변환 및 가공하고, 마지막으로 현재 상용화되어 있는 머신 러닝 기법들을 개별 논변의 분석 및 코딩에 적용시키고 그 결과가 연구자의 코딩과 비교할 때 얼마나 높은 일치도를 얻을 수 있는지를 보고한다.

II. 이론적 배경

1. 국내 문헌에서의 논변 분석 방법

타당한 논변을 구성하여 설득력을 높이기 위해 논변(argument)은 주장을 뒷받침하는 다양한 요소들로 구성된다. 이때 주장을 어떠한 체계로 뒷받침하였는지, 즉 논변의 구조를 분석하기 위해 Toulmin (1958)이 삼단논법을 대신하여 고안한 틀이 논변의 구조를 드러낸 대표적인 틀로서 활용되고 있다. Toulmin은 논변을 구성하는 요소들로 주장(claim), 보장(warrant), 자료(data), 보강(backing), 반박(rebuttal), 한정어(qualifier)를 들었다. 구체적으로 살펴보면, 먼저 주장은 논변 활동에서 다른 사람을 설득시키려고 하는 것이다. 그리고 자료는 왜 그 주장이 타당한지를 뒷받침하기 위한 토대이며, 보장은 자료가 주장을 어떻게 뒷받침하는지 보여주고, 보강은 보장이 왜 타당한지 뒷받침한다. 한정어는 논변이 성립하는 조건을 제시함으로써 주장에 대한 확신의 정도를 나타내며, 반박은 주장이 성립되지 않는 조건에 해당한다. Toulmin의 논변 구조 틀은 그 실용성으로 인해 다양한 분야에서 널리 활용되고 있으며, 국내 과학 교육 분야에서 또한 많은 연구들이 Toulmin의 논변 구조 틀을 활용하여 학생들이 구성한 논변의 구조를 분석하고 있다(e.g. Han *et al.*, 2012; Lee, Yun, & Kim, 2015).

하지만 과학적 논변의 영역 특이적 성격으로 인해, Toulmin(1958)의 틀이 과학 수업에서 이루어지는 학생들의 논변 활동에 잘 부합하지 않으며 학생의 논변 활동 자료에서 어느 부분이 논변의 각 구성 요소에 해당하는지 분석하기 어렵다는 점이 보고되어왔다(Jiménez-Aleixandre, Rodríguez, & Duschl, 2000; Sampson & Clark, 2008). 이에 많은 연구에서 논변 활동에서 이루어진 학생 담화를 바탕으로 Toulmin의 틀을 변형하여 논변의 구조를 분석하였다(e.g. Han *et al.*, 2012; Lee, Lee, & Kim, 2015). 또한 학생의 발화 또는 학생이 작성한 문장이 논변의 구성 요소 중 어느 것에 해당하는지가 모호한 경우가 많아, 자료 분석 단계에서 여러 연구자들이 각자 코딩을 거친 후에 코딩 결과를 서로 대조해보고, 서로 코딩 결과가 다른 경우에는 합의 도출하는 과정을 통해 분석의 타당성을 확보하고 있다.

2. 한국어 자연어 처리

과학 교실에서 이루어지는 학생들의 논변 활동을 분석하기 위해서는 실제 교실 환경에서 녹음된 발화를 텍스트로 옮기고 이를 컴퓨터 프로그램에서 사용하기에 적절한 자료구조로 변환하는 작업이 요구된다. 음성 인식(Speech Recognition) 및 이를 텍스트로 변환하는 작업(Speech To Text, STT)은 머신 러닝 분야의 주요 연구 주제 중 하나이지만, 기술 개발이 주로 영미권에서 이루어져 아직까지는 한국어 분석하는 데 있어 그 기능이 상당히 제한적이다(Choi, & Kim, 2017). 그러므로 현재까지는 여전히 연구자가 직접 전사본을 작성해야 하는 상황이다.

한편 발화를 컴퓨터가 분석하기 위해서는 인간 발화를 컴퓨터가 처리할 수 있는 형태로 변환하는 자연어 처리(Natural Language Processing, NLP)가 요구된다. 최근 한글 텍스트를 분석하기 위한 자연어 처리에 요구되는 코드들을 구조화하여 묶어둔 패키지 또한 다양하게 개발되어 보급되고 있다. 그 대표적인 사례로 통계 계산을 위한 프로그래밍 언어인 R에서 사용 가능한 ‘KoNLP’ 패키지(Jeon, 2016) 및 객체지향적 프로그래밍 언어의 일종인 Python에서 사용 가능한 ‘KoNLPy’ 패키지(Park, & Cho, 2014)가 있는데, 여기서는 본 연구에서 사용한 후자를 주로 살펴보고자 한다. KoNLPy는 기존에 다양한 프로그래밍 언어로 개발되어 있던 한나눔(Hannanum), 꼬꼬마(Kkma), Komoran, McCab, Twitter 등의 형태소 분석기를 내부 모듈로 포함할 뿐 아니라 한국 법률 말뭉치 및 대한민국 국회 의안 말뭉치 등의 데이터, 그리고 KAIST 말뭉치 또는 세종 말뭉치를 이용해 생성된 사전 등을 제공함으로써 한국어 자연어 처리를 위한 종합적인 기능을 제공하는 패키지이다(Park, & Cho, 2014). 예컨대, 입력한 텍스트에서 용례를 검색하거나, 형태소를 분석하고 품사를 태깅(tagging)하거나, 최빈 형태소를 찾는 작업이 비교적 쉽고 빠른 프로그래밍 언어인 Python 내에서 이루어질 수 있도록 구현되어 있다. 이렇듯 자연어 처리를 수월하게 해 주는 패키지들이 공개됨에 따라, 교육학적 맥락에서도 이들을 활용하여 실과 교과의 핵심 개념을 분석(Go, 2016)하거나 2015 개정 교육과정에 대한 언론 보도를 분석하는 등(Yu, & Baek, 2016)의 연구가 이루어지기도 하였다.

여기서 관련 선행 문헌들이 분석하고자 하는 대상과 관련한 단어들을 자연어 처리 패키지의 내부 사전에 등록하였음에 유의하여야 한다(e.g. Yu, & Baek, 2016). 아무리 뛰어난 패키지라 하더라도 연구자의 관심 분야에 속하는 모든 용어를 포함하기는 어렵기 때문이다. 예컨대, KoNLPy의 꼬꼬마 모듈의 사전을 구축하는데 활용된 세종 말뭉치의 경우 총 1,294개의 말뭉치 파일 중 약 102개 정도가 자연과학과 연관이 있는 것으로 분류되고 있다. 다른 분야에 비하여 빈약한 과학 분야의 탑재 용어수로 미루어 보아, 과학 교과에서의 논변 활동을

보다 정확하게 분석하기 위해서는 과학 관련 용어들을 형태소 분석 모듈의 내부 사전에 추가해야 할 필요가 있음을 알 수 있다.

한편, 연구를 진행하는 과정에서 어떠한 형태소(또는 단어)들을 취사선택하여 추가적으로 분석할 것인지에 대한 기준 또한 하나의 쟁점이 될 수 있다. 전체 문서에서 특정한 언어 단위가 출현하는 빈도(Term Frequency, TF)를 살펴보는 방법도 선행 연구들에서 많이 사용하였던 방법이었으나 이는 단순히 해당 언어에서 많이 사용되기 때문일 뿐 개별 문서들의 특성을 나타내지는 못한다는 비판을 받기도 한다(Lee, Kim, & Jung, 2016). 이에 근래의 연구에서는 특정 언어 단위의 TF값에 해당 언어 단위가 출현하는 문서 빈도의 역수(Inverse Document Frequency, IDF)를 곱해준 TF-IDF값을 활용하고 있다. 예컨대, 어떤 언어 단위의 TF값이 높게 나타나더라도 그 언어 단위가 거의 모든 문서들에서 나타난다면 IDF값이 작게 되므로 TF-IDF 값이 떨어지게 되며, 결과적으로 그 중요성이 낮다는 의미로 이해할 수 있다. 만약 TF값이 높으면서 IDF값 또한 높다면, 이는 해당 언어 단위가 특정한 문서에서만 많이 출현한다는 의미이므로 문서의 특성을 이해하는데 중요한 역할을 할 것으로 기대할 수 있다. 여기서 IDF값을 계산하는 방법은 여러 가지가 있지만, 일반적으로는 특정 언어 단위가 출현하는 문서의 빈도(Document Frequency, DF)값의 역수에 로그를 취하는 경우가 많다.

3. 텍스트 분류를 위한 머신 러닝 기법

머신 러닝은 텍스트, 음성, 이미지 등을 여러 자료를 분류하거나, 수치적으로 예측하거나, 패턴을 탐지하거나, 군집화하는 등 다양한 용도로 사용된다. 머신 러닝은 일반적으로 학습(훈련) 데이터를 통해 모델을 구축하고, 기존에 주어지지 않았던 테스트 데이터에 관한 일반화가 얼마나 잘 이루어졌는지를 평가하는 순서로 이루어진다. 이때 학습 데이터조차 잘 설명하지 못하는 과소적합(underfitting)이 일어나서도 안 되지만, 학습 데이터에만 담긴 패턴을 포착하여 다른 변수들 간의 관계를 잘못 유추하는 과적합(overfitting)이 일어나는 것도 바람직하지 않다. 이러한 머신 러닝 알고리즘은 학습 데이터를 토대로 새로이 주어진 테스트 데이터에 대한 예측값을 산출하기 위한 예측 모델(predictive model)을 구축하는 지도 학습기(supervised learner)와 기술 모델(descriptive model)을 구축하는 자율 학습기(unsupervised learner)로 구분된다. 과학 논변 활동을 코딩하는 것은 학습 데이터를 기반으로 생성된 모델이 테스트 데이터가 어떻게 분류될 지를 예측하는 것이므로 본 연구에서는 지도 학습기를 주로 살펴 보았다. 텍스트 분류기로 활용될 수 있는 머신 러닝 기법들의 원리와 특징을 살펴보자면 다음과 같다(Ivezić *et al.*, 2014; Lantz, 2013) (Table 1).

Table 1. A Summary of the Properties of Some Machine Learning Classifiers (Ivezić *et al.*, 2014; Lantz, 2013)

방법	정확성	해석가능성	단순성	속도	관련 R 패키지
k-최근접 이웃 (KNN)	높음	높음	높음	중간	“class” (Venables, & Ripley, 2002)
나이브 베이즈 분류 (NBC)	낮음	높음	높음	높음	“e1071” (Meyer <i>et al.</i> , 2017)
결정 트리 학습 (DTL)	중간	높음	높음	중간	“tree” (Ripley, 2016)
인공신경망 (ANN)	높음	낮음	낮음	중간	“neuralnet” (Fritsch, & Guenther, 2016)
(커널형) 서포트 벡터 머신 (SVM)	높음	낮음	낮음	낮음	“kernlab” (Karatzoglou <i>et al.</i> , 2004)

k-최근접 이웃 알고리즘(k-Nearest Neighbors algorithm, KNN)은 범주를 알지 못하는 예제를 이미 분류되어 있는 가장 유사한 예제의 범주에 속하는 것으로 지정한다. 즉, 일반적으로 최근접 분류기는 동질적인 수치를 지닌 사례들을 동일 범주로 지정한다. 여기서 k는 범주화해야 할 사례에 대하여 가장 가까운 k개의 사례들이 어느 범주에 속하였는지를 살펴보고, 그 중 다수인 범주를 택한다는 것을 의미한다. 나이브 베이즈 분류(Naïve Bayes Classifier, NBC) 기법은 베이시안(bayesian) 통계학의 아이디어로부터 출발한 것으로서, 특정한 단어가 포함되어 있는지의 여부에 따라 텍스트를 분류하는 데 주로 사용된다. 이러한 나이브 베이즈 기법은 단순하고 빠르며 효과적이므로 사실상 문서 분류의 표준에 해당하며, 실제로 활용되는 대표적인 분야로는 스팸 메일 분류가 있다. 결정 트리 학습(Decision Tree Learning, DTL)은 이른바 ‘나누어 정복하기(divide and conquer)’식의 접근으로 논리 결정을 위한 트리 구조를 만드는 것을 의미한다. 전체 데이터를 대표하는 뿌리 노드(node)에서 시작하여, 주어진 범주화와 가장 가까운 결과를 내는 결정 노드(decision node)들과 가지(branch, vertex)들을 자동적으로 생성한다. 인공신경망(Artificial Neural Network, ANN)은 생물체의 뉴런이 신호를 입력받고 출력하는 과정을 머신 러닝에 적용한 기법이다. 인공신경망을 이루는 여러 개의 인공 뉴런들은 다양한 수와 층으로 이루어진 망을 이루면서 각자가 받은 입력 신호를 활성 함수에 의하여 출력 신호로 바꾸고 다음 인공 뉴런에 전달하는 역할을 한다. 사용자가 인공신경망의 구조를 정의하고 데이터를 통해 훈련을 반복하면 각 뉴런들 간의 연결 가중치를 변화시키는 알고리즘이 작동하면서 모델을 향상시키게 되며, 이러한 인공신경망에서 은닉 층의 수를 늘려나가는 형태가 인공지능 기술을 획기적으로 발전시킨 것으로 평가받는 딥 러닝 기술의 기반이다(Zhang, 2017). 서포트 벡터 머신(Support Vector Machine, SVM)은 동질적인 데이터들이 놓일 수 있는 초평면(hyperplane)을 만드는 수학적 기법이다. 여기서 서포트 벡터란 두 범주에 속하는 점들을 나누는 여러 평면 중 최대 마진 초평면(Maximum Margin Hyperplane, MMH)을 선택할 때 그에 가장 가까운 점들의 벡터를 의미한다. 서포트 벡터 머신의 특징 중 하나는 비선형 공간에서도 커널(kernel)이라고 알려진 방법을 통해 문제를 고차원 공간으로 확장시켜 보다 쉽게 초평면을 찾을 수 있다는 점이다.

위와 같은 머신 러닝 기법들은 문제와 데이터의 특성에 따라 최적의 결과를 낼 수 있는 조건이 다르다. 그러므로 사용자가 머신 러닝을 활용하고자 하는 분야에 있어서 어떤 기법을 사용하는 것이 가장 좋은지에 관한 탐색과 시행착오가 요구된다.

III. 연구 방법

1. 과학 논변 활동 데이터 수집

국내 문헌에서 Toulmin의 논변 구조 틀(1958)을 바탕으로 학생들이 구성한 논변의 구조가 어떻게 분석해 왔는지 검토하기 위해 먼저 학술지에 게재된 논문들을 검색하였다. 학술지에 게재된 논문들의 경우 그 분석의 질을 보장받을 수 있을 것이라고 판단하여, 학술지에 게재된 논문들로 검색 범위를 제한하였다. 논변(argument)은 논증, 논의 등 한국어에서 여러 용어로 번역되기에, 한국교육학술정보원

(KERIS)에서 제공하는 학술연구정보서비스인 RISS(www.riss.kr)에서 ‘논변’, ‘논증’, ‘논의’를 검색어로 하여 문헌을 검색하였다. 이 중에서 과학 수업에서 이루어지는 논변 활동을 다룬 연구들을 찾기 위해 ‘과학’을 재검색하여 검색 범위를 좁혔으며, 한국과학교육학회지, 생물교육, 대한화학회지, 대한지구과학교육학회지, 교과교육연구, 현장과학교육, 초등과학교육 등 과학 교육과 관련된 학술지에 게재된 문헌을 재검색하였다. 그리고 검색된 문헌들의 내용을 검토하여, 논변 담화 또는 글쓰기에서 Toulmin의 논변 구조 틀에 따라 논변 구조를 코딩한 결과가 연구 결과에서 제시된 문헌 총 18건을 도출하였다. 18건의 문헌에서 추출된 논변은 총 990개 문장이었으며, 이를 머신러닝을 위한 학습 데이터로 사용하였다.

이 때 본 연구에서는 논변의 구조를 머신 러닝을 통해 분석한 결과와 사람이 분석한 결과 사이의 일치도를 확인하고자 하였으므로, 앞서 도출한 각 문헌의 자료 분석 단계에서 논변 구조를 어떻게 분석하였는지를 탐색하였다. 문헌에서 학생의 논변 활동을 분석한 방법은 크게 두 갈래로 나뉘었다(참고자료 1). 7개의 문헌은 Toulmin의 틀을 따랐으나 각 구성 요소의 의미를 재정의하거나 동료 연구자들과 합의를 도출하는 과정을 거쳤으며, 11개의 문헌은 분석한 자료로부터 명확히 드러나지 않은 구성 요소를 제외하고 구성 요소의 의미를 수정하였다. 일관된 기준에 따라 논변의 구성 요소를 코딩한 결과를 바탕으로 머신 러닝에 쓰일 데이터와 그 산출물과의 일치도를 확인하기 위해, Table 2과 같이 각 논변 구성 요소를 재정의하였다. 이러한 정의는 과학 교육에서 Toulmin(1958)의 틀의 적합성에 대한 검토를 바탕으로 새로 도출된 McNeill & Krajcik(2007)의 틀, 그리고 앞서 검토한 국내 문헌들을 바탕으로 하여 도출되었다(e.g., Han *et al.*, 2012; Kwon & Kim, 2016; Lee *et al.*, 2015). 그리고 이 기준에 따라 기존 문헌에 제시된 자료를 재코딩하였고, 이때 논변의 구성 요소에 해당되지 않는 발화는 NULL로 표기하였다. 논변 구성 요소의 재정의와 코딩 결과의 타당성을 높이기 위해 연구자 2인이 각자 코딩한 후, 코딩 결과에 있어서 합의를 도출하였다.

또한 실제 교실 환경에서 발화된 논변을 머신 러닝의 테스트 데이터로 사용하고, ‘소집단 학생들의 논변활동을 지원하기 위한 교수 학습 전략 개발’ 및 ‘과학적 논변활동에서 학생들의 프레이밍 이해를 통한 프레이밍-반응적 교수·학습 전략 구축’ 프로젝트에서 소집단 논변 활동을 학교 현장에 도입하여 얻은 자료를 활용하였다. 이 프로젝트들에는 경기도 수원 소재 중학교 1학년 한 학급의 학생 37명과 담당 교사 1명, 서울시 소재의 중학교 2학년 한 학급의 학생 30명과 담당 교사 1명이 참여하였다. 학생들은 3~4명으로 구성된 소집단으로 나뉘어 논변 활동에 참여하였다. 중학교 1학년을 대상으로 설계된 논변 활동은 ‘광합성’ 단원에, 그리고 중학교 2학년을 대상으로 설계된 논변 활동은 ‘자극과 반응’ 단원에 해당하는 학습 내용에 관한 것이었다. 논변 활동에서의 학생들의 담화를 수집하기 위해 각 소집단 별 논의 과정을 녹음·녹화하였으며, 논변 활동이 활발하게 이루어진 소집단의 담화를 전사하였다. 앞서 재정의한 논변의 구성 요소 분석 틀을 바탕으로 이 담화본에서 드러나는 학생들의 발화를 분석하여, 마찬가지로 연구자들이 각자 코딩한 뒤 코딩 결과에 있어서 합의를 도출하였다. 여기서 추출된 논변은 총 483개 문장이었다.

이후 각 출처에서 논변의 구조를 분석한 논변 활동 자료와 연구자들의 코딩 결과를 정리하였다. 출처마다 각 990개와 483개의 문장을

기법의 머신 러닝을 수행하였다. 이는 Table 1에 제시된 바와 같은 R 패키지들을 활용하여 이루어졌으며, 각각의 기법에서 최적의 정확도를 낼 수 있기 위하여 개별 기법에서 필요로 하는 변수의 수치를 변화시키며 여러 조건에서의 머신 러닝을 수행하였다(예: k-최근접 이웃법에서의 k값). 이후, 테스트 데이터에서 머신 러닝이 생성한 모델의 코딩 결과와 연구자의 코딩 결과가 얼마나 유사한지를 살펴보기 위하여 R의 'caret' 패키지에서 제공하는 confusionMatrix() 함수를 활용해 둘 간의 혼동 행렬(confusion matrix), 일치도(accuracy), Kappa(κ) 등을 구하였다. 다만, 해당 조건에서 모델링을 위한 계산이 수렴하지 않거나 심각한 과소적합 및 과적합이 일어났다고 판단된 경우에는 가능한대로 조건을 달리하여 다시 머신 러닝을 수행하였다. 예컨대, 인공신경망(ANN)의 경우 표면 노드에서 20개를 초과하는 형태소를 입력으로 하거나 은닉 노드의 수와 층을 늘릴 경우 계산 수렴 속도 및 일치도가 감소하는 경향을 보였으므로 불가피하게 상위 20개의 형태소만을 입력으로 하였으며 10개의 은닉 노드를 포함하는 1개의 은닉 층만을 설정하였다.

논변-형태소-품사 행렬의 성격을 생각해 볼 때, 머신 러닝으로 논변 1개 문장을 코딩하기 위하여 제공한 정보는 기본적으로 해당 문장 안에 어떠한 형태소가 포함되었는지에 해당한다. 이를테면 논변의 전후 맥락을 고려하지 않더라도 단순히 문장 1개만 고려하였을 때 적절한 코딩이 이루어질 수 있는지를 살펴본 것이다. 이러한 방법으로 머신 러닝을 수행한 이후, 논변의 맥락을 고려하는 것이 코딩의 정확도에 얼마나 영향을 미치는지 알아보기 위하여 바로 직전에 발화된 문장이 연구자 혹은 머신 러닝의 결과로 생성된 모델에 의하여 어떻게 코딩되었는지에 관한 정보만을 요인(factor) 형태로 추가하여 다시 한 번 머신 러닝을 수행하였다. 담화의 맥락을 고려하는 여러 방법 중 직전 발화를 염두에 두는 것이 가장 간단한 형태라고 할 수 있기 때문이다(Kim, 2003). 이 때, 각 맥락의 첫 문장에 해당하는 발화에 대하여는 직전 문장의 코딩을 '미분류(NULL)'로 처리하였다.

4. 개발 환경

데이터 전처리와 머신 러닝을 위한 프로그램 개발과 실행은 연구자 1인의 노트북 PC에서 수행하였다. 운영체제는 Windows 10이었으며 하드웨어 환경은 i7-7500U CPU와 16GB RAM이었다. 프로그램 개발은 Python 3.6.3과 PyCharm Community Edition 2017.2.4. 및 R 3.4.2와 R Studio 1.1.383에서 이루어졌다. 프로그램 개발 및 머신 러닝 과정에서 병렬 처리는 고려하지 않았으며, GPU 등 병렬 처리를 도울 수 있는 장치 또한 사용하지 않았다.

IV. 연구 결과

1. 수집된 논변 분석 자료에 관한 기술 통계

본 연구에서 수집한 1,473개의 과학 논변들 중 학습 데이터로 사용한 990개 문장은 기존에 보고된 선행 연구로부터 인용 및 재코딩한 것이었으며, 테스트 데이터로 사용한 483개 문장은 연구자들이 전사 데이터로부터 코딩한 것이었다. 이러한 논변들이 본 연구에서 재구성된 Toulmin의 틀에 따라 분류된 비율은 Table 3과 같았다.

학습 데이터와 테스트 데이터 모두에서 논변들은 '미분류(NULL)', '정당화', '주장', '자료', '반박'의 순으로 많이 코딩되어 있었다. 특히, 미분류된 문장이 가장 많았다는 점은 교사가 지식을 전달하는 전통적인 교수학습 방식에 익숙한 학생들이 과학 논변 활동에서 논변을 구성해나가는 생산적인 실행에 잘 참여하지 못한다는 선행 연구(Berland & Hammer, 2012)와 다양한 상호작용이 이루어지는 실제 교실 현장의 모습이 반영된 것으로 보인다. 또한 '반박'이 양 쪽 모두에서 10% 미만이었던 점은 학생들이 논변 활동 과정에서 반박을 잘 수행하지 못한다는 선행 연구와 일치한다(e.g. Osborne, Erduran, & Simon, 2004; Shin & Kim, 2011).

한편 학습 데이터와 테스트 데이터의 차이도 존재하였다. 그 중 가장 두드러지는 것은 미분류된 문장의 비율이라고 볼 수 있는데, 학습 데이터에서는 미분류된 문장이 약 27% 정도였으나 테스트 데이터에서는 약 48% 정도였다. 그 외에 '반박' 및 '정당화'의 비율은 양 쪽에서 비슷하였으나, '자료'나 '주장'은 테스트 데이터에서 다소 낮은 비율을 차지하는 것으로 나타났다. 이러한 차이는 테스트 데이터는 실제 교실 환경에서 이루어진 논변의 전사본을 코딩한 결과 전체를 사용한 것임에 비하여, 학습 데이터는 선행 논문에서 추출한 것이기 때문에 발생한 것으로 보인다. 특히 학술 논문의 경우 미분류된 문장들을 어느 정도 제외하고 유의미한 형태로 코딩된 논변들을 보다 많이 보고하였기 때문에 미분류된 문장의 비율이 상대적으로 낮게 나타났을 수 있다.

Table 3. Descriptive Statistics of Coded Arguments

구분	미분류 (NULL)	자료	주장	반박	정당화	전체
학습 데이터	266 (26.87%)	177 (17.88%)	225 (22.73%)	81 (8.18%)	241 (24.34%)	990 (100%)
테스트 데이터	232 (48.03%)	51 (10.6%)	60 (12.42%)	34 (7.04%)	106 (21.94%)	483 (100%)
계	498 (33.81%)	228 (15.48%)	285 (19.34%)	115 (7.81%)	347 (23.56%)	1,473 (100%)

2. 논변 자료 처리 결과

분석 결과 1,473개의 문장에서 추출된 형태소는 총 2,077종 23,339개로 문장 1개당 평균 11.09개의 형태소가 포함되어 있었다. 형태소들의 종류와 빈도 및 출현 비율을 품사에 따라 정리한 결과는 Table 4와 같다. 이를 통해 학생들의 과학 논변 활동에서 형태소의 품사는 체언(25.88%), 어미(21.52%), 조사(16.64%), 용언(15.78%) 순으로 많이 나타남을 알 수 있다. 여기서 총 10종 28개의 형태소가 '분석 불능'으로 분류되었는데, 이는 'KoNLPy' 패키지 내의 사전에 등록되지 않은 '애네들', '재' 등 학생들의 자연스러운 발화 과정에서 나타나는 단어들이므로 나타났다. 이러한 분석 불능 단어들은 약 0.1%에 불과하였다.

한편, 추출된 2,077종의 형태소들의 전체 출현 빈도(TF)를 보정하여 개별 형태소가 특정한 문장들에서 많이 나타날수록 값이 크게 나타나는 TF-IDF값의 합이 400 미만인 형태소 및 기타 문장 부호 등을 제외한 결과 42종의 형태소가 남게 되었다. 여기서 논변의 주제에 특수적으로 나타나는 것으로 판단되는 명사 '물'은 제외하고 41종의

Table 4. Descriptive Statistics of Morphemes

	체인	용언	관형사	부사	감탄사	조사	어미	접사	어근	부호	한글 이외	분석 불능	전체
종류	1,047	427	20	154	34	78	195	21	26	9	56	10	2,077
빈도	6,040 (25.88%)	3,682 (15.78%)	294 (1.26%)	1,049 (4.5%)	240 (1.03%)	3,883 (16.64%)	5,022 (21.52%)	673 (2.88%)	83 (0.36%)	2,240 (9.6%)	105 (0.45%)	28 (0.12%)	23,339 (100%)

Table 5. Morphemes Used to Machine Learning

품사	형태소	세부 품사 (Kkma 태그)	빈도(TF)	TF-IDF	품사	형태소	세부 품사 (Kkma 태그)	빈도(TF)	TF-IDF
체인 (4개)	것	의존 명사 (nmb)	217	655.06	어미 (14개)	다	평서형 종결 어미 (efn)	472	885.98
	거	의존 명사 (nmb)	169	560.25		는	관형형 전성 어미 (etd)	356	873.30
	여기	대명사 (np)	103	438.55		ㄴ	관형형 전성 어미 (etd)	297	814.73
	수	의존 명사 (nmb)	105	418.47		지	보조적 연결 어미 (ecs)	262	698.64
용언 (7개)	이	긍정 지정사 (vcp)	366	817.29		고	대등 연결 어미 (ece)	222	673.26
	하	동사 (vv)	259	726.09		게	보조적 연결 어미 (ecs)	186	616.61
	있	동사 (vv)	219	666.41		었	시제 선어말 어미 (ept)	184	599.43
	않	보조 형용사 (vxa)	140	493.44		ㄴ	시제 선어말 어미 (ept)	194	596.23
	되	동사 (vv)	118	454.55		아	의문형 종결 어미 (efq)	181	574.82
	아니	부정 지정사 (vcn)	113	430.56		면	의존적 연결 어미 (ecd)	171	561.93
	않	보조 동사 (vxv)	111	421.43		ㄷ	관형형 전성 어미 (etd)	165	556.84
부사 (1개)	더	접속 부사 (mag)	125	479.76		어	평서형 종결 어미 (efn)	161	541.71
조사 (14개)	이	주격 조사 (jks)	677	1073.42		어	보조적 연결 어미 (ecs)	105	418.47
	에	부사격 조사 (jkm)	411	957.234		어서	의존적 연결 어미 (ecd)	101	407.30
	가	주격 조사 (jks)	482	945.83	접미사 (1개)	하	동사 파생 접미사 (xsv)	395	916.39
	을	목적격 조사 (jko)	295	811.18					
	는	보조사 (jx)	317	800.28					
	의	관형격 조사 (jkg)	224	701.62					
	은	보조사 (jx)	233	679.71					
	를	목적격 조사 (jko)	140	503.13					
	에서	부사격 조사 (jkm)	124	475.92					
	으로	부사격 조사 (jkm)	123	466.98					
	로	부사격 조사 (jkm)	118	461.36					
	도	보조사 (jx)	116	460.51					
	요	보조사 (jx)	118	456.54					
	고	관형격 조사 (jkq)	114	440.76					

형태소를 선정한 결과는 Table 5와 같으며, 이에 따라 생성한 논변-형태소-품사 행렬은 Figure 2와 같다. Figure 2에 나타난 행렬의 의미를 예를 들어 설명하자면, 1번 문장에서 주격 조사 ‘이(jks)’가 2번 나타났으므로(TF), 이를 전체 1,473개 문장에서 ‘이(jks)’가 나타난 횟수의 역수의 로그값(IDF)으로 나누어준 TF-IDF값 3.17이 셀의 값으로 나타나 있다. 한편 동사 파생 접미사 ‘하(xsv)’는 1번 문장에서 나타나지 않으므로 셀의 값이 0이다. Table 5를 Table 4와 비교할 때 체언과 용언의 종류가 상당히 줄어들고 조사와 어미는 상대적으로 많은 종류가 잔류하였음을 알 수 있는데, 이는 체언과 용언의 경우 논변의 주제에 따라 매우 다양한 종류가 작은 빈도로 나타날 수 있기 때문으로 보인다. 예컨대, 빛을 주제로 하는 논변에서는 ‘빛’(체인)이 ‘비추다’(용언)는 형태소가 사용될 수 있지만 지진을 주제로 하는 논변에서는 ‘지진’(체인), ‘땅’(체인), ‘흔들리다’(용언) 등 전혀 다른 형태소가 사용된다. 그에 비하여 ‘~이’, ‘~에서’, ‘~으로’ 등의 조사나

‘~는’, ‘~면’, ‘~어서’ 등의 어미는 논변의 주제와 관계 없이 다양한 상황에서 나타날 수 있는 형태소이다. 결국, 이러한 결과는 과학 논변 활동을 분석함에 있어 체언과 용언보다는 조사나 어미에 집중하는 것이 보다 효율적일 수 있음을 암시한다.

3. 텍스트 분류기 머신 러닝 결과

학습 데이터를 활용하여 머신 러닝을 수행하고 테스트 데이터에 관한 예측 코딩을 실시한 결과, 테스트 데이터에 관한 연구자의 코딩과의 일치도는 Table 6~Table 8과 같이 나타났다. 먼저 Table 6은 테스트 데이터의 개별 문장의 형태소에 관한 정보만을 제공한 결과이다. 여러 모델들의 예측 결과는 연구자의 코딩 결과와 비교할 때 모두 45% 이상의 일치도를 보였으며, κ 또한 0.2 이상인 것으로 나타났다. 이 때 가장 높은 성능을 보인 모델은 k-최근접 이웃 알고리즘(KNN)으

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1		0 jks	에 jkm	가 jks	하 xsv	다 efn	는 etd	이 vcp	ㄴ etd	을 jko	는 jx	하 vv	의 jkg	지 ecs	은 jx	고 ece	있 vv	것 nnb	게 ecs
2	1	3.175808	2.334717	1.962299	0	0	0	0	2.743182	0	2.524542	0	6.264449	0	0	0	6.113872	3.032689	0
3	2	1.587904	0	1.962299	0	0	0	0	0	0	0	0	0	0	2.917211	0	3.056936	0	0
4	3	3.175808	4.669434	7.849197	0	0	2.453079	2.239139	0	0	5.049083	5.606885	6.264449	0	2.917211	0	0	0	0
5	4	1.587904	0	1.962299	0	0	0	0	0	2.749755	0	0	0	2.666561	0	0	0	0	0
6	5	1.587904	0	0	0	0	0	0	0	0	5.049083	0	0	2.666561	0	0	0	0	0
7	6	0	0	0	0	0	0	4.478279	0	0	0	0	3.132224	0	0	0	0	0	0
8	7	0	2.334717	1.962299	0	0	0	0	2.743182	2.749755	2.524542	5.606885	3.132224	0	0	3.032689	0	0	0
9	8	1.587904	2.334717	0	0	0	0	0	2.743182	0	2.524542	0	3.132224	2.666561	0	3.032689	0	0	0
10	9	4.763711	0	0	0	0	2.453079	0	0	0	2.524542	0	9.396673	2.666561	0	0	0	0	0
11	10	1.587904	0	3.924599	0	0	0	0	5.486364	0	5.049083	0	3.132224	2.666561	0	0	0	0	0
12	11	1.587904	0	0	0	0	0	0	2.743182	0	0	0	0	0	0	0	0	0	0
13	12	0	0	1.962299	0	0	0	2.239139	0	0	0	0	0	0	0	0	0	0	0
14	13	0	0	0	2.319971	0	0	2.239139	0	0	0	0	0	0	0	0	0	0	0
15	14	1.587904	0	1.962299	0	0	2.453079	0	0	0	0	0	0	0	0	0	0	3.032689	0
16	15	1.587904	0	0	2.319971	0	0	0	2.743182	0	0	0	0	0	0	0	0	0	3.315088
17	16	0	0	0	0	0	0	0	0	0	2.524542	2.803443	0	0	0	3.032689	0	0	0
18	17	1.587904	0	1.962299	0	0	0	0	2.743182	0	2.524542	0	0	0	0	0	0	0	0
19	18	0	0	0	0	1.877083	0	0	0	0	0	2.803443	0	0	2.917211	3.032689	0	0	0
20	19	1.587904	0	0	0	0	2.453079	0	0	0	0	0	0	0	0	0	0	0	0

Figure 2. Argument – Morpheme:Class Matrix

로 54.04%의 일치도와 $\kappa = 0.22$ 의 결과를 도출하였으며, 그 다음으로는 서포트 벡터 머신(SVM)이 53.62%의 일치도와 $\kappa = 0.32$ 의 결과를 도출하였다. 이는 학생들의 담화에서 맥락을 고려하지 않는 코딩에서 서포트 벡터 머신이 가장 높은 정확도를 보인다는 선행 연구(Rosé et al., 2008)와도 대체로 일치한다. 한편 이는 개별 문장의 형태소만 살펴보는 경우에도 절반 이상의 정확도로 학생들의 과학 논변을 코딩할 수 있음을 보여준다.

한편 Table 7과 Table 8은 학습 과정에서 다른 조건들은 동일하게 유지하되 직전 문장이 어떻게 코딩되었는지에 관한 정보를 추가적으로 제공한 결과이다. 결정 트리 학습(DTL)을 제외하고는 머신 러닝 결과 이전과 다른 예측 모델들이 생성되었으며, 인공신경망(ANN)의 경우 이전 문장에 대한 정보를 요인(factor) 형태로 제공하였으므로 표면 노드의 수가 증가하였다. Table 7의 경우 개별 문장의 형태소뿐만 아니라 그 직전 문장을 연구자들이 어떻게 코딩하였는지에 관한 정보를 요인의 형태로 고려하도록 한 결과로서, 이 중 k-최근접 이웃 알고리즘(KNN)이 55.07%로 가장 높은 일치도와 $\kappa = 0.24$ 의 결과를 도출하였으며, 서포트 벡터 머신(SVM) 또한 54.87%의 일치도와 $\kappa = 0.34$ 로 가장 높은 값을 도출하였다. 이를 Table 6과 비교할 경우, 직전 문장의 코딩 정보를 제공하여 새로 생성된 모델들이 미세하나마 보다 나은 일치도와 κ 값을 보이는 경향을 확인할 수 있다. 이러한 결과는 Rosé et al., (2008)이 인공신경망(ANN)과 서포트 벡터 머신(SVM)을 활용하여 논변의 구성 요소에 대한 미소 수준(micro level)의 분석을 진행한 결과 논변 활동이 일어나는 맥락을 고려하는 것이 머신 러닝의 성능 향상에 약간의 긍정적 효과를 가져다주는데 그친다고 보고하였던 것과 일치하는 것으로서, 한국어로 이루어진 과학 논변 활동을 여러 머신 러닝 기법으로 분석한 결과가 영어 사용 환경에서 이루어진 선행 연구와 유사하게 나타났다는 점에 주목할 만하다.

한편 Table 8은 Table 7에서와 동일한 예측 모델을 사용하였지만 테스트 데이터의 직전 문장에 관한 정보로 연구자의 코딩이 아니라 머신 러닝에 의해 예측된 코딩을 순차적으로 제공하는 과정을 반복한 결과이다. 이를 Table 7과 비교해보면 예측 결과의 일치도 및 κ 값이 대체로 감소하였음을 알 수 있다. 이는 직전 문장에 대하여 연구자의 코딩이 아닌 예측 모델에 의한 코딩이 정보로 주어질 경우 발생하는 오차가 개입하였기 때문으로 이해할 수 있다. 한편 Table 8을 Table 6과 비교해보면, 대체로 Table 8에서의 경우가 일치도가 낮은 가운데

k-최근접 이웃 알고리즘(KNN)은 일치도 및 κ 값이 Table 6의 경우와 비교하여 상대적으로 높게 나타났다. 이는 직전 문장의 코딩 결과에 대한 정보로 인하여 일치도가 상승하는 효과와 예측 모델의 오차로 인하여 일치도가 감소하는 효과가 서로 상충하여 발생한 현상으로 짐작된다. 예컨대, k-최근접 이웃 알고리즘(KNN)에서는 둘 중 전자의 효과가 더 크게 나타나 일치도 혹은 κ 값이 상승하였고, 나머지 경우에는 후자의 효과가 더 크게 나타나 일치도가 감소하였을 수 있다.

Table 6. Comparison of Machine Learning Models - Considering Each Sentence Respectively

	KNN	NBC	DTL	ANN	SVM
일치도	54.04%	52.59%	50.93%	45.13%	53.62%
Kappa (κ)	0.22	0.27	0.24	0.22	0.32
기타 설정	k = 41	-	가지 수 = 4	표면 노드 수 = 20, 은닉 노드 수 = 10×1	kernal = "splinedot"

Table 7. Comparison of Machine Learning Models - Considering Real Code of Previous Sentence

	KNN	NBC	DTL	ANN	SVM
일치도	55.07%	53.62%	50.93%	47.20%	54.87%
Kappa (κ)	0.24	0.3	0.24	0.25	0.34
기타 설정	k = 41	-	가지 수 = 4	표면 노드 수 = 20+5, 은닉 노드 수 = 10×1	kernal = "splinedot"

Table 8. Comparison of Machine Learning Models - Considering Predicted Code of Previous Sentence

	KNN	NBC	DTL	ANN	SVM
일치도	54.66%	52.17%	50.93%	44.54%	52.38%
Kappa (κ)	0.23	0.27	0.24	0.23	0.30
기타 설정	k = 41	-	가지 수 = 4	표면 노드 수 = 20+5, 은닉 노드 수 = 10×1	kernal = "splinedot"

4. 머신 러닝 결과의 해석

위에서와 같은 머신 러닝 결과는 개별 문장들이 포함된 형태소 정보만으로도 연구자의 코딩 결과와 약 55%가 일치하는 자동화된 코딩을 수행할 수 있음을 보여주었다. 하지만 이러한 결과를 해석함으로써 더 많은 함의를 도출하는 작업 또한 필요하다.

먼저, 직전 문장에 대한 정보를 제공하였을 때 일치도 및 κ 값이 증가한 정도가 상대적으로 높았던 인공신경망(ANN)의 혼동 행렬을 비교했을 때, 직전 문장에 관한 정보 유무가 논변 코딩에 어떠한 영향을 미치는지를 다음 세 가지 측면에서 직간접적으로 살펴볼 수 있다(Tables 9, 10). 첫째, 직전 문장의 코딩 정보는 현재 문장이 'NULL'에 해당하는지 혹은 논변의 구성 요소 중 하나로서 올바르게 코딩하는데 긍정적인 효과를 보였다. 이는 특정 발화가 논변의 구성 요소로서 의미를 갖는지의 여부가 해당 상황 맥락에 의해 영향을 받는다는 점이 반영된 결과로서 이해 가능하다. 둘째, 논변의 구성 요소 중 일치도의 상승폭이 비교적 높았던 '주장'의 경우는 직전 문장이 '정당화'로 분류된 경우에 더욱 정확하게 예측이 이루어졌다. 이는 Table 3과 Table 11을 비교했을 때 학습 및 테스트 데이터 전반에서 '정당화'는 약 24%이지만 '주장'으로 코딩된 논변 직전의 문장들이 '정당화'로 코딩된 비율은 약 42%로 비교적 높다는 점과, 모델의 수치적 의미를 이해하기 쉽지 않은 인공신경망(ANN) 대신 Table 7의 나이브 베이즈 분류(NBC)의 결과에서 직전 문장이 '정당화'로 코딩된 경우에는 현재 문장이 '주장'으로 코딩되도록 하는 가중치가 약 0.43으로서 다른 선택지들의 가중치 0.18~0.23에 비해 2배가량 높다는 점에서 뒷받침된다. 셋째, 직전 문장의 정보를 고려하였을 때 논변의 구성 요소 중 '자료' 및 '주장'의 분류에 있어 정확도가 다소 상승하였으며, Table 9에서는 '반박'으로 예측된 문장이 전혀 없었던 반면 Table 10에서는 비록 올바른 예측은 아니었을지라도 '반박'으로 예측된 경우가 나타났다. 이는 데이터 상에서 본래 '반박' 및 '자료'의 비율이 가장 적었기 때문에 발생하였던 과적합의 문제가 직전 문장의 코딩 정보에 의해 어느 정도 해소될

Table 9. Confusion Matrix of ANN - Considering Each Sentence Respectively

예측	NULL	반박	자료	정당화	주장
NULL	135	11	12	17	18
반박	0	0	0	0	0
자료	11	8	7	18	8
정당화	52	12	27	60	19
주장	34	3	5	11	15

Table 10. Confusion Matrix of ANN - Considering Real Code of Previous Sentence

예측	NULL	반박	자료	정당화	주장
NULL	148	9	13	22	20
반박	2	0	0	2	0
자료	25	8	11	13	9
정당화	26	11	13	50	9
주장	31	6	14	19	22

Table 11. Frequency of Codes of Sentences before 'Claim' Sentences

NULL	반박	자료	정당화	주장	전체
65 (22.81%)	20 (7.02%)	46 (16.14%)	121 (42.46%)	33 (11.58%)	285 (100%)

수 있음을 보여준다. 위의 세 측면과는 달리 직전 문장의 코딩 정보를 반영한 경우인 Table 10에서 개별 문장만을 고려한 Table 9에 비하여 '정당화'를 올바르게 분류한 사례가 60개에서 50개로 다소 감소하였다는 점도 관찰할 수 있으나, 그만큼 다른 문장들을 '정당화'로 잘못 분류한 사례 또한 110개에서 59개로 크게 감소하였음을 감안하여야 한다.

그런가 하면, 머신 러닝 기법들 중 상대적으로 시각적인 표현 및 그 해석이 용이한 결정 트리 학습(DTL)에 의하여 생성된 모델을 살펴볼 때 논변 코딩에 어떤 형태소들이 큰 영향을 미치는 지를 보다 쉽게 파악할 수 있다. Figure 3은 머신 러닝에 의해 생성된 결정 트리를 빈도에 따른 순서도와 같이 나타낸 것으로서 Table 6~8에서와 같이 약 51%의 일치도를 보인 모델이다. 이 모델에서는 평서형 종결 어미 '다(efn)', 부사격 조사 '에(jkm)', 의문형 종결 어미 '아(efq)'가 등장하는지의 여부만을 화살표와 로 나타낸 순서대로 고려하여 어떤 논변이 'NULL', '자료', '주장', '정당화' 중 무엇으로 코딩될지를 결정한다. 예컨대, 이러한 간단한 결정 트리를 따라서도 테스트 데이터 중 "저희 의견은, 타당하지 않다." 라는 논변은 올바르게 '주장'으로, "거기에는 기공은 안보이는 거지?"는 '정당화'로, "이런 경우는 되잖아. 그거 뭐지. 차타고 가면서, 어, 소리를 듣잖아, 이어폰을 끼면은"은 '자료'로, "그런데 그렇다 해도 왜 애네가 애네가 뭔지 궁금해"는 'NULL'로 올바르게 분류되었다. 이러한 결과는 '주장'에서는 다른 부가적인 의미를 전달하기보다는 단정적인 어조로 정보 전달만을 주요 기능으로 하는 평서형 종결 어미 "-다"가 주로 사용된다는 점(Lim, 2011), '정당화'에서는 데이터를 바탕으로 한 추론, 즉 학생들이 주목한 정보를 바탕으로 새로운 의견을 제시하는 내용이 담겨있으므로 특정 최소나 시간을 지목하거나 비교 대상 간의 차이를 말하고자 할 때 주로 사용되는 부사격 조사인 '-에'(Zhao, 2016)가 주로 등장한다는 점, '자료'에서는 화자가 정보의 내용을 이미 알고 있다는 전제 하에 그 정보를 알리거나 청자에게 확인을 요청할 때 구어체 '~잖아' 등과 같이 의문형 종결 어미 '아(efq)'가 포함된 표현이 등장한다는 점(Won, 2011)을 밝힌 선행 연구 문헌들로 뒷받침된다. 하지만 Figure 3과 같은 모델은 그 혼동 행렬을 살펴볼 때 '반박'이 가장 적은 비율로 나타나던 학습 데이터에 대한 과적합이 일어나, 결과적으로 테스트 데이터를 '반박'으로 분류할 가능성이 고려되지 않은 지나치게 단순화된 모델이다. 여기서 학습 데이터 상 '반박'의 비율을 높이기 위해 학습 데이터의 크기를 900문장 정도로 축소하여 과적합을 방지한 결과 Figure 4에서와 같이 '반박'으로의 분류 가능성을 포함하는 보다 일반적인 모델을 얻을 수 있었다. 약 48%의 일치도를 보이는 Figure 4의 모델에 따르면, 평서형 종결 어미 '다(efn)'가 포함되어 있으면서 부정 지칭사 '아니(vcn)'를 포함하는 문장은 '반박'에 해당할 가능성이 높다. 예컨대, 다소 문어적인 표현이지만 "아니다."라는 표현을 포함하는 논변은 '반박'에 해당할 개연성이 충분하다.

한편 Figure 5는 Table 6에서와 같이 약 45%의 일치도를 보이는

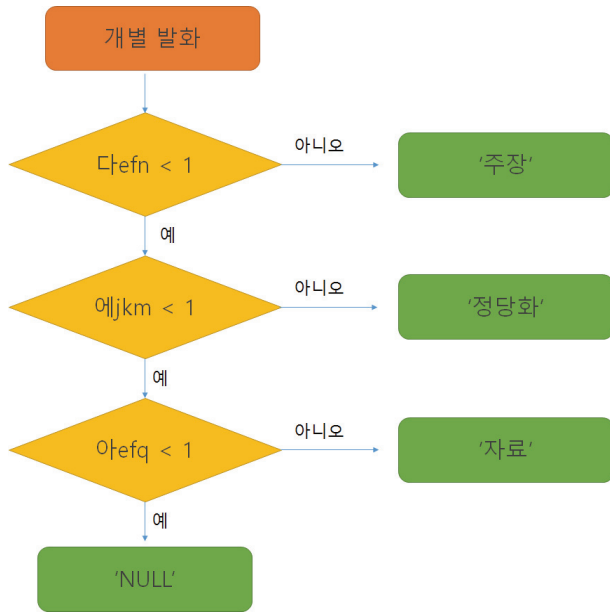


Figure 3. Plot of DTL with Original Training Data

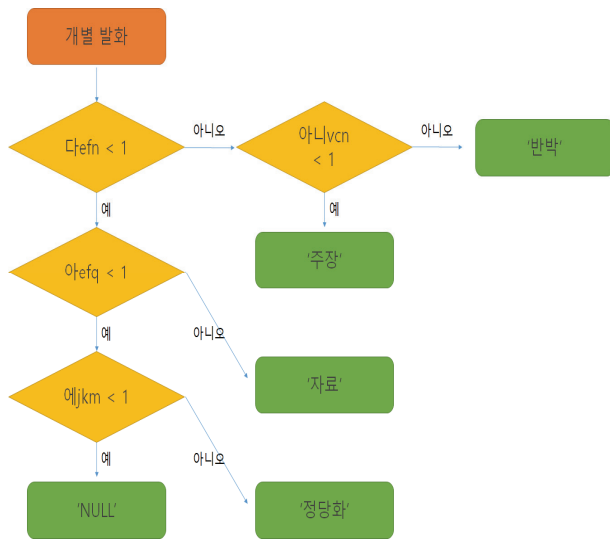


Figure 4. Plot of DTL with Reduced Training Data

모델을 인공신경망(ANN) 모델을 가시화한 결과이다. 여기서 맨 왼쪽에는 20개의 형태소에 상응하는 표면 노드들로 이루어진 표면 층이, 가운데에는 10개의 은닉 노드로 이루어진 1개의 은닉 층이, 맨 오른쪽에는 5개의 논변 유형에 상응하는 출력노드들로 이루어진 출력 층이 나타나 있다. 입력 층의 노드들에는 개별 문장들의 형태소 별 TF-IDF 값들(Figure 2)이 입력으로 들어오게 되고, 머신 러닝 결과로서 노드들 각각에는 다음 층의 각 노드들로 향하는 연결 가중치가 계산되어 있으므로 표면 층과 은닉 층 사이의 200개의 연결과 은닉 층과 출력 층 사이의 50개의 연결, 곧 250개의 연결 관계를 통과하여 출력 층의 5개 노드 각각이 결과값을 내놓게 된다. 출력 층의 노드 중에서 가장 큰 값을 지니는 것으로 그 문장을 분류하는 것이다. 이처럼 인공신경망(ANN)의 경우에는 각 노드들 간의 관계가 수치적으로 나타남에도, 이를 연구자들이 이해할 만한 이론적 형태로 나타내기가 어렵다는 점에서 일종의 ‘블랙박스(blackbox)’와 같은 모델을 형성한다는 단점을 지니고 있다(Lantz, 2013). 이와 같은 인공신경망에서 표면 층과 은닉 층의 합이 3 이상일 때를 딥 러닝(deep learning)의 일종이라고

할 수 있으며, 은닉 층 및 은닉 노드의 수가 많아질 때 신경망 내에서 정보가 전달될 수 있는 경로가 복잡해지고 다양해지므로 보다 높은 정확도를 갖는 모델을 생성할 가능성이 있다. 다만 딥 러닝의 경우 은닉 층을 더 깊게 형성해야 할 뿐 아니라 과도한 데이터 전처리를 하지 않고 최대한 많은 데이터를 사용하여야 한다(Zhang, 2017). 예컨대, 추후 연구에서 더욱 깊은(deep) 인공신경망(ANN)을 활용하여 더 나은 일치도를 보이는 모델을 생성하고자 하는 경우에는 본 연구에서와 같이 TF-IDF 값이 높은 일부 형태소만을 사용할 것이 아니라 전체 2,000여 종의 형태소들 중 더 많은 부분을 학습 데이터로 사용해야 할 것이다. 이 때 증가하게 되는 계산 시간 또한 염두에 두어야 하며, 연구자가 더욱 이해하기 어려운 블랙박스가 되더라도 일치도가 높은 모델을 추구할 것인지 혹은 일치도가 상대적으로 낮더라도 해석이 용이한 모델을 추구할 것인지를 고찰해보아야 한다.

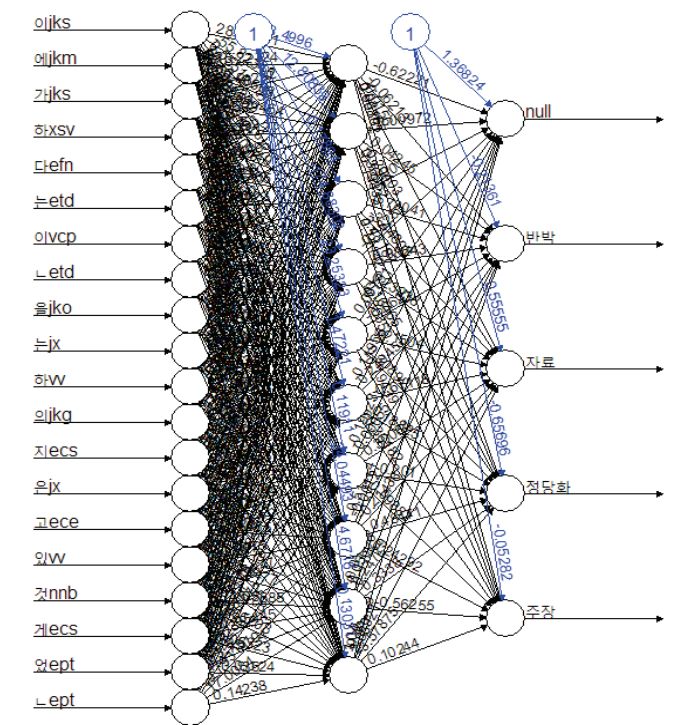


Figure 5. Plot of ANN in Table 6

이러한 요소들을 종합적으로 고려해볼 때, 본 연구의 머신 러닝 결과는 다음과 같은 세 가지 함의를 지닌다고 할 수 있다. 첫째로, 머신 러닝을 통한 논변의 구성 요소 분석이 개별 문장과 같이 기초적인 수준에서도 이루어질 수 있다는 점을 보였다. 실제로 머신 러닝 과정에서 개별 문장의 형태소만을 고려하여도 54% 이상의 일치도를 보이는 모델을 생성할 수 있었으며, 결정 트리를 시각화하여 어떠한 형태소들이 이러한 분류에 큰 영향을 미치는지를 알아볼 수 있었다. 둘째로, 과학 논변을 분석할 때 직전 문장과 같이 그 맥락을 고려해야 함을 보였다. 특히 인공신경망(ANN)에 의해 생성된 혼동 행렬의 비교를 통해, 맥락에 대한 고려가 특정 발화가 논변의 구성 요소에 해당 하는지를 구분하고, 논변 구성 요소 내에서는 어디에 해당하는지를 판별하며, 모델의 학습 데이터에 대한 과적합을 해소하는데 도움이 될 수 있음을 제시하였다. 셋째로, 탐색적으로 이루어진 본 연구의 방법을 발전시킨다면 더욱 우수한 모델을 생성할 수 있다는 가능성을

보였다. 예컨대, 본 연구에서는 답화의 맥락을 고려하는 일환으로서 직전 문장의 코딩 정보만 제공하였지만 이보다 심화된 방법을 사용할 수 있을 것이며, 본 연구에서와는 다른 방법으로 형태소를 분석 및 선정하거나, 더욱 발전된 형태의 머신 러닝이라고 할 수 있는 딥 러닝 기법을 활용할 수 있을 것이다. 또한 본 연구는 과학적 논변을 분석함에 있어서 기본적으로 머신 러닝에 의하여 데이터로부터 자동 생성된 모델을 고찰하였지만, 추후에는 여기에 연구자가 기존에 알고 있었던 중요한 요인들을 포함시키는 작업도 뒤따라야 할 것이다.

V. 결론 및 논의

본 연구에서는 국내 교육학 연구에서 거의 사용되지 않던 머신 러닝 기술을 과학 교육 연구에 접목하여, 학생들의 과학 논변 활동에서 나타나는 논변의 구성 요소를 분석하는 과정을 자동화할 수 있는 가능성을 탐색해보았다. 학습 데이터로는 Toulmin이 제안하였던 틀에 따라 학생들의 과학 논변 구성 요소를 코딩한 국내 선행 문헌 18건을 수합하고 정리하여 990개의 문장을 추출하였으며, 테스트 데이터로는 실제 교실 환경에서 발화된 과학 논변 전사 데이터를 사용하여 483개의 문장을 추출하고 연구자들이 사전 코딩을 수행하였다. Python의 'KoNLPy' 패키지와 '꼬꼬마(Kkma)' 모듈을 사용한 한국어 자연어 처리(Natural Language Processing, NLP)를 통해 개별 논변을 구성하는 단어와 형태소를 분석하였으며, 연구자 2인과 국어교육 석사학위 소지자 1인의 검토 과정을 거쳤다. 총 1,473개의 문장에 대한 논변-형태소-품사 행렬을 만든 후에 다섯 가지 방법으로 머신 러닝을 수행하고 생성된 예측 모델과 연구자의 사전 코딩을 비교한 결과, 개별 문장의 형태소만을 고려하였을 때에는 k-최근접 이웃 알고리즘(KNN)이 약 54%의 일치도($\kappa = 0.22$)를 보임으로써 가장 우수하였다. 직전 문장이 어떻게 코딩되어 있는지에 관한 정보가 주어졌을 때, k-최근접 이웃 알고리즘(KNN)이 약 55%의 일치도($\kappa = 0.24$)를 보였으며 다른 머신 러닝 기법에서도 전반적으로 일치도가 상승하였다. 본 연구의 결과는 과학 논변 활동의 분석에서 개별 문장을 고려하는 단순한 방법이 어느 정도 유용함과 동시에, 답화의 맥락을 고려하는 것 또한 필요함을 데이터에 기반하여 보여준다. 또한 머신 러닝을 통해 교실에서 한국어로 이루어진 과학 논변 활동을 분석하여 연구자와 교사들에게 유용하게 사용될 수 있는 가능성을 보여준다.

최근 과학 교육 분야에서 과학 공동체의 인식적 실행이 강조되면서, 학생들의 학습 산출물뿐만 아니라 교수학습 과정과 그 상호작용의 특성을 포착하려는 연구들이 이루어지고 있다. 교수학습 과정 분석에 머신 러닝을 접목시키는 것은 이러한 작업에 소요되는 노력과 시간을 절감하는 등의 장점을 통해(Luckin *et al.*, 2016) 과학 교육 분야의 연구와 교실 현장에서의 실천이 변화하는 일에 상당한 기여를 하게 될 것이다. 본 연구의 경우는 머신 러닝 기법을 과학 교수학습 과정의 분석에 적용하여 얼마만큼의 일치도를 얻을 수 있는지를 탐색하였다. 이는 인공지능 기술이 교육 연구에 도입되기 시작하는 시점에 국내 과학 교육 맥락에서의 그 가능성과 발전 방향을 제안하였다는 점에서 의의를 지닌다.

구체적으로 살펴보자면, 첫째로 본 연구는 학생들의 논변 활동 담화로부터 논변의 구성 요소를 분석하는 방법을 다루며, 머신 러닝을 활용한 분석을 바탕으로 기존 분석 방식을 되돌아보는 계기를 제공한

다. 기존 문헌에서는 과학 수업에서의 논변 활동에서 학생들이 논변을 구성해 나가는 과정을 포착하기 위해 주로 Toulmin(1958)의 틀을 활용하였다. 하지만 각 연구에서 분석하는 논변 활동의 맥락에 대한 연구자의 판단에 따라 논변의 구조나 각 구성 요소의 의미가 서로 다르게 설명되어왔다. 본 연구는 결정 트리 학습(DTL)과 같은 머신 러닝을 활용하여 형태소의 출현 빈도와 같이 언어의 형태론적 측면을 기반으로 논변의 구조를 분석하였다. 그리고 이를 통해 논변의 구성 요소별로 등장하는 종결 어미와 조사 등에 주목할 필요성을 데이터에 기반하여 실증적으로 제안하였다. 이처럼 기존 문헌에서 논변 활동을 분석할 때 주목하지 않았던 측면을 고려함으로써, 본 연구는 논변의 구조 분석에 있어 객관성과 타당성을 확보하기 위한 또 다른 방안의 가능성을 시사한다. 추후 학습 데이터가 더 많이 축적된다면, 머신 러닝을 통한 논변의 구조 분석이 연구 방법의 타당성을 높이는 또 다른 방안으로서 활용될 수 있을 것이다. 이와 같은 맥락에서 머신 러닝을 활용한 학생들의 상호작용 분석은 과학 교육에서의 교수학습 연구 방법론의 발전을 촉진할 것으로 기대되며, 본 연구는 그 초석이 될 수 있다는 점에서 시사점을 지닌다.

둘째로 본 연구는 머신 러닝을 통하여 한국어로 이루어지는 논변 활동을 분석할 수 있는 방안을 모색하여, 국내 교수학습 환경에서의 논변 활동을 지원하는 전략을 발전시키는 기반을 마련하였다. 전통적인 교수학습에서보다 논변 활동에서 교사는 학생들의 논의로부터 수많은 정보를 접하고 해석하는 과정에서 많은 인지적 부담을 안는다(McNeill, 2009). 주로 학생들의 기록물을 통해 학습 과정을 파악하던 전통적인 교수학습 방법에 비해 논변 활동에서는 말을 통한 대화적 상호작용이 강조되기 때문에, 교사에게는 수업 현장에서의 순간적인 인지와 판단 능력이 요구되는 것이다(Nielson, 2013). 교수 실행에서의 이와 같은 어려움을 지원하기 위하여 교수학습 전략과 자료가 만들어져왔으나(e.g. Sampson & Schleigh, 2013; McNeill, 2009), 학생들의 논변 활동으로부터 새로이 접하는 정보를 자동으로 분석할 수 있도록 지원하는 방안은 거의 연구되지 않았다. 특히 한국어 자연어 처리를 다루는 지원 방안은 드물었다. 이러한 측면에서 많은 양의 데이터를 빠른 속도로 자동 처리하기 위하여 머신 러닝을 활용한 학생들의 논변 활동 분석 방법을 탐색한 본 연구는 교실 현장에서 교사의 인지적 부담을 경감해 줄 방안(Pardo, 2014)을 모색하였다는 점에서 그 의의를 지닌다.

셋째로 본 연구의 결과는 추후 머신 러닝을 과학 교육의 다른 활동 맥락에서도 활용하기 위한 연구의 초석이 될 것이다. 예컨대, 본 연구에서는 과학 교실에서 이루어진 학생들의 발화를 분석 대상으로 삼았지만 외국에서 이루어진 선행 연구처럼 온라인상의 논변과 같은 글쓰기 활동을 분석하는 일에도 머신 러닝을 사용할 수 있을 것이다(e.g. Goodman *et al.*, 2005; Soller, 2004). 더 나아가 이는 학생들이 온라인 글쓰기에서 논변의 구성 요소를 고려하도록 자동으로 지원하는 방안을 탐색하는 연구로 이어질 수 있을 것이다. 또한 기존의 머신 러닝이 SNS 기록으로부터 감성적 측면을 분석하는 데에도 활용된 사례가 많다는 점에서(e.g. Go, Bhayani, & Huang, 2009), 학생들의 발화와 감성적 측면 간의 관계를 탐색하는 연구로도 이어질 수 있다. 이처럼 본 연구는 머신 러닝 기술의 과학 교육에의 도입 가능성을 살핍으로써, 과학 교수학습 지원 전략 및 탐색 분야를 확장시킬 수 있는 기반을 마련하였다는 점에서 그 시사점을 지닌다.

다만 본 연구는 머신 러닝을 교육학 연구에 접목하기 위한 탐색적 성격을 지니므로 그 제한점 또한 지니고 있다. 기술적 측면에서 현재로서는 한국어 음성 인식 오픈 API의 성능이 충분하지 않은 상황이다. 이에 본 연구에서는 교실에서 일어나는 학습 과정에 대한 동시적인(synchronous) 분석의 가능성을 염두에 두고 수행되었지만 불가피하게 전사본 데이터를 활용한 비동시적(asynchronous) 학습 분석을 진행하였다. 또한 본 연구에서 사용한 자료구조인 논변-형태소품사 행렬은 셀의 값이 0이 많은 희소행렬(sparse matrix)에 해당하여 정보의 양이 적었고(Figure 2), 기계가 텍스트의 뜻을 이해하지 못하는 ‘얕은 텍스트 처리(shallow text processing)’를 수행하였다. 다음으로는 이론적인 측면이다. 본 연구에서는 연구자의 견해보다는 데이터에 기반하여 기계가 자동으로 모델을 생성한다는 머신 러닝의 특성(Domingos, 2015; Lantz, 2013; Lim, 2015; Zhang, 2017)을 되도록 유지하였다. 추후에는 데이터에서 곧바로 추출하기는 쉽지 않지만 연구자가 이미 알고 있는 중요한 이론적 요소들을 포함하는 후속 연구가 이루어져야 하며, 이를 통해 연구자들이 모델을 더욱 깊이 이해하고 적용하도록 해야 한다. 마지막으로 데이터의 측면이다. 본 연구에서는 머신 러닝 결과 논변의 구성 요소에 해당하는 것과 그렇지 않은 것(NULL)의 구분은 잘 일어나지만 논변의 구성 요소 간의 분류는 보다 개선이 필요함을, 그리고 학생들의 논변 중 가장 비율이 적은 ‘반박’에 대한 과(소)적합이 일어남을 관찰하였다. 여기서 머신 러닝의 특성상 양질의 데이터가 다량으로 확보 가능할 경우 더 나은 예측 모델을 구축할 수 있음을 염두에 두어야 한다. 특히, 교실 환경에서 이루어지는 학생 간의 상호작용이 구체적으로 이루어진다는 점에 유의하여야 할 것이다.

본 연구의 결과와 제한점을 바탕으로, 후속 연구 및 이를 위한 기반 구축을 다음과 같이 제안하고자 한다. 먼저 본 연구와 같은 기술적 한계를 극복하기 위해서는 연구 목적으로 활용 가능한 수준의 한국어 음성 인식 기술이 개발 및 공개되어야 할 필요가 있다. 또한 머신 러닝을 보다 다양한 논변 활동에 적용하기 위해서는 실제 교실 환경에서 발화된 논변 데이터가 풍부하게 수집되고 공유되어야 한다. 예컨대, 충분한 양의 데이터가 확보된다면 논변 활동에서 주로 등장하는 형태소들이 잘못 분류되지 않도록 분석 프로그램의 사전에 미리 등록시켜 보다 정확한 분석이 이루어지도록 하는데 도움이 될 수 있다. 이처럼 본 연구의 제한점을 보완함으로써 본 연구에서 사용한 바와 같은 머신 러닝 기법들을 발전시켜 각각에 관한 모델을 구축하고 개선하며 해석하는 일이 요구된다(Luckin *et al.*, 2016). 이를 기반으로 더 나은 머신 러닝 모델을 구축할 뿐 아니라 연구자들이 익숙한 이론적 요소를 개입시킬 수 있을 것이며, 이러한 모델을 바탕으로 구체적인 논변 활동의 맥락 속에서 이루어지는 학생들의 인식적 실행에 관한 이해를 심화할 수 있을 것이다. 이러한 후속 연구들이 충분히 이루어진 뒤에는, 궁극적으로 논변 활동 코딩 프로그램을 스마트 기기로 인식 및 구동하여 실제 교실 환경에서 학생들의 발화를 동시적으로 인식하고 이를 즉각적으로 분석하며 그 결과를 교사와 학생에게 피드백해 주는 형태의 학습 분석 및 관리 시스템을 구축하는 교육공학 적 작업을 시도할 수 있을 것이다. 특히, 학생들의 담화와 그 분석 자료를 기록으로 저장할 수 있다는 점을 활용함으로써, 학생과 교사가 자신의 실행을 반성할 수 있도록 지원하는 등 교수학습 과정의 추가적 지원을 모색하는 연구로 확장될 수 있을 것으로 기대한다.

국문요약

본 연구에서는 국내 교육학 연구에서 거의 사용되지 않던 머신 러닝 기술을 과학 교육 연구에 접목하여, 학생들의 과학 논변 활동에서 나타나는 논변의 구성 요소를 분석하는 과정을 자동화할 수 있는 가능성을 탐색해보았다. 학습 데이터로는 Toulmin이 제안하였던 틀에 따라 학생들의 과학 논변 구성 요소를 코딩한 국내 선행 문헌 18건을 수합하고 정리하여 990개의 문장을 추출하였으며, 테스트 데이터로는 실제 교실 환경에서 발화된 과학 논변 전사 데이터를 사용하여 483개의 문장을 추출하고 연구자들이 사전 코딩을 수행하였다. Python의 ‘KoNLPy’ 패키지와 ‘꼬꼬마(Kkma)’ 모듈을 사용한 한국어 자연어 처리(Natural Language Processing, NLP)를 통해 개별 논변을 구성하는 단어와 형태소를 분석하였으며, 연구자 2인과 국어교육 석사학위 소지자 1인의 검토 과정을 거쳤다. 총 1,473개의 문장에 대한 논변-형태소품사 행렬을 만든 후에 다섯 가지 방법으로 머신 러닝을 수행하고 생성된 예측 모델과 연구자의 사전 코딩을 비교한 결과, 개별 문장의 형태소만을 고려하였을 때에는 k-최근접 이웃 알고리즘(KNN)이 약 54%의 일치도($\kappa = 0.22$)를 보임으로써 가장 우수하였다. 직전 문장이 어떻게 코딩되어 있는지에 관한 정보가 주어졌을 때, k-최근접 이웃 알고리즘(KNN)이 약 55%의 일치도($\kappa = 0.24$)를 보였으며 다른 머신 러닝 기법에서도 전반적으로 일치도가 상승하였다. 더 나아가, 본 연구의 결과는 과학 논변 활동의 분석에서 개별 문장을 고려하는 단순한 방법이 어느 정도 유용함과 동시에, 담화의 맥락을 고려하는 것 또한 필요함을 데이터에 기반하여 보여주었다. 또한 머신 러닝을 통해 교실에서 한국어로 이루어진 과학 논변 활동을 분석하여 연구자와 교사들에게 유용하게 사용될 수 있는 가능성을 보여준다.

주제어 : 과학 논변, 머신 러닝, 인공지능, 자동화, 자연어 처리

References¹⁾

- Baek, J., Jeong, D. H., & Hwang, S. (2014). Issues and effects in developing inquiry-based argumentation task for science teachers: A case of Charles' law experiment. *Journal of the Korean Association for Science Education*, 34(2), 79-92.
- Berland, L. K., & Hammer, D. (2012). Framing for scientific argumentation. *Journal of Research in Science Teaching*, 49(1), 68-94.
- Booth, W. C., Colomb, G. G., & Williams, J. M. (2008). *The craft of research*. Chicago: The University of Chicago Press.
- Cho, H. -A., Chang, J. -E., & Kim, H. -B. (2013). Epistemic level in middle school students' small-group argumentation using first-hand or second-hand data. *Journal of the Korean Association for Science Education*, 33(2), 486-500.
- Choi, S. J., & Kim, J. -B. (2017). Comparison analysis of speech recognition open APIs' accuracy. *Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology*, 7(8), 411-418.
- Domingos, P. (2015). *The master algorithm: How the quest for ultimate learning machine will remake our world*. NY: Basic Books.
- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84(3), 287-312.
- Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research*

1) ‘neuralnet’, ‘e1071’, ‘tree’, ‘kernlab’ 등의 일부 R 패키지의 경우 그 이름이 개발자들이 명명한 고유명사에 해당하며 이를 대문자로 인용할 경우에는 오히려 독자들이 원 자료에 도달하기가 어려워진다고 판단하여 불가피하게 소문자로 시작하도록 인용하였다.

- in Education, 32(1), 268-291.
- Erduran, S., Simon, S., & Osborne, J. (2004). TAPing into argumentation: Developments in the application of Toulmin's argument pattern for studying science discourse. *Science Education*, 88(6), 915-933.
- Fritsch, S., & Guenther, F. (2016). *neuralnet: Training of neural networks*. R package version 1.33. Retrieved January 3, 2018, from <https://CRAN.R-project.org/package=neuralnet>
- Go, A., Bhayani, R., & Huang, L. (2009). Twitter sentiment classification using distant supervision. CS224N Project Report, Stanford.
- Go, I. -K. (2016). The key concept study of agriculture unit in practical arts based on big data analysis methods. *Journal of Learner-Centered Curriculum and Instruction*, 16(4), 431-450.
- Goodman, B. A., Linton, F. N., Gaimari, R. D., Hitzeman, J. M., Ross, H. J., & Zarrella, G. (2005). Using dialogue features to predict trouble during collaborative learning. *User Modeling and User-Adapted Interaction*, 15(1), 85-134.
- Han, H. -J., Lee, T., Ko, H., Lee, S. -K., Kim, E., Choe, S. -U., & Kim, C. -J. (2012). An analysis of the type of rebuttal in argumentation among science-gifted student. *Journal of the Korean Association for Science Education*, 32(4), 717-728.
- Heo, H., Yang, E. -J., Kim, D., Choi, J. G., & Moon, Y. S. (2017). Human intelligence and learning in the era of artificial intelligence. *The Korean Journal of Philosophy of Education*, 39(1), 101-132.
- Ivezić, Ž., Connolly, A. J., VanderPlas, J. T., & Gray, A. (2014). *Statistics, data mining, and machine learning in astronomy: A practical Python guide for the analysis of survey data*. NJ: Princeton University Press.
- Jang, G. -H. (2003). A trial on reconstruction of written language/spoken language and their styles. *Korean Semantics*, 13, 143-165.
- Jeon, H. (2016). *KoNLP: Korean NLP package*. R package version 0.80.1. Retrieved January 3, 2018, from <https://CRAN.R-project.org/package=KoNLP>
- Jiménez-Aleixandre, M., Rodríguez, A. B., & Duschl, R. A. (2000). "Doing the lesson" or "doing science": Argument in high school genetics. *Science Education*, 84(6), 757-792.
- Jung, C. Y., & Seo, Y. H. (1998). Artificial intelligence: machine translation of Korean-to-English spoken language based on semantic patterns. *The KIPS Transaction*, 5(9), 2361-2368.
- Karatzoglou A., Smola A., Hornik, K., & Zeileis, A. (2004). *kernlab - An S4 package for kernel methods in R*. *Journal of Statistical Software* 11(9), 1-20. Retrieved January 3, 2018, from <http://www.jstatsoft.org/v11/i09/>
- Kim, I. -H. (2003). The nature of context and its relations to language use. *English Language Education*, 26, 7-34.
- Kim, K., & Park, Y. (2017). Development and application of the teaching and learning model of artificial intelligence education for elementary students. *Journal of The Korean Association of Information Education*, 21(1), 137-147.
- Kwon, J. -S., & Kim, H. -B. (2016). Exploring small group argumentation shown in designing an experiment: Focusing on students' epistemic goals and epistemic considerations for activities. *Journal of the Korean Association for Science Education*, 36(1), 45-61.
- Lantz, B. (2013). *Machine learning with R*. Packt Publishing Ltd.
- Larrain, A., Freire, P., & Howe, C. (2014). Science teaching and argumentation: One-sided versus dialectical argumentation in Chilean middle-school science lessons. *International Journal of Science Education*, 36(6), 1017-1036.
- Lederman, N. G. (1992). Students' and teachers' conceptions of the nature of science: A review of the research. *Journal of Research in Science Teaching*, 29, 331-359.
- Lee, D., Yeon, J., Hwang, I., & Lee, S. (2010). KKMA: A tool for utilizing sejong corpus based on relational database. *Journal of KIISE: Computing Practices and Letters*, 16(11), 1046-1050.
- Lee, E. J., Yun, S. M., & Kim, H. -B. (2015). Exploring small group argumentation and epistemological framing of gifted science students as revealed by the analysis of their responses to anomalous data. *Journal of the Korean Association for Science Education*, 35(3), 419-429.
- Lee, J., & Kim, H. -B. (2011). Small group argumentation pattern of middle school students constructed in the conflict context. *Biology Education*, 39(2), 235-247.
- Lee, J., Nam, C., & Shin, D. (2017). Machine learning bigdata education platform using apache spark. *Proceedings of the Korea Software Congress 2017*.
- Lee, J. -A., Maeng, S. -H., & Kim, C. -J. (2008). A new way of reading the science classroom discourse: Pedagogical discourse analysis. *Journal of the Korean Association for Science Education*, 28(8), 832-847.
- Lee, J. -S., Kang, K. -H., & Lee, S. -K. (2005). Students' interactions of reasoning and inductions in small group science discussion. *The Journal of the Smeieccu*, 26(1), 91-112.
- Lee, K. H., Yun, S. -M., & Kim, H. -B. (2012). Understanding of middle school students' small group argumentation of plant and animal classification: Focusing on the effects of leader. *Biology Education*, 40(10), 71-86.
- Lee, S., Kim, H. -J., & Jung, S. (2016). Text-mining based topic analysis on online sexism. *Journal of Cybercommunication Academic Society* 33(3), 159-199.
- Lee, S., Lee, C., & Lee, G. (2010). Example-based dialog system for english conversation tutoring. *Journal of KISS : Software and Applications*, 37(2), 129-136.
- Lee, S., Park, S. -H., & Kim, H. -B. (2016). Exploring secondary students' progression in group norms and argumentation competency through collaborative reflection about small group argumentation. *Journal of the Korean Association for Science Education*, 36(6), 895-910.
- Lee, S. -K. (2006). The patterns and the characteristics of students' interactive argumentation in the small-group discussions. *Journal of the Korean Chemical Society*, 50(1), 79-88.
- Lee, Y., Lee, S., & Kim, H. -B. (2015). Understanding students' knowledge construction and scientific argumentation according to the level of openness in inquiry and the abstraction level of scientific knowledge. *Biology Education*, 43(1), 50-69.
- Lee, Y. -O. (2002). Translation problems between Korean and English reflecting their structural differences: With respect to the translation of reported speech. *The Journal of Translation Studies*, 3(1), 59-81.
- Lim, D. (2011). Sentence types in Korean. *Journal of Korean Linguistics*, 60(4), 323-359.
- Lim, D. (2015). *Bigdata analysis using R*. Paju: Freeacademy.
- Lim, J., Song, Y. -M., Song, M., & Yang, I. -H. (2010). An analysis on the level of elementary gifted students' argumentation in scientific inquiry. *Journal of Korean Elementary Science Education*, 29(4), 441-450.
- Lim, H., & Yeo, S. -I. (2012). Characteristics on elementary students' argumentation in science problem solving process. *Journal of Korean Elementary Science Education*, 31(1), 13-24.
- Luckin, R., Holmes, W., Griffiths, M., & Forcier, L. B. (2016). *Intelligence unleashed: An argument for AI in education*. London: Pearson.
- Maeng, S., Park, Y. -S., & Kim, C. -J. (2013). Methodological review of the research on argumentative discourse focused on analyzing collaborative construction and epistemic enactments of argumentation. *Journal of the Korean Association for Science Education*, 33(4), 840-862.
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2017). *e1071: Misc functions of the department of statistics, probability theory group (formerly: E1071), TU Wien*. R package version 1.6-8. Retrieved January 3, 2018, from <https://CRAN.R-project.org/package=e1071>
- McAlister, S., Allen, C., Ravenscroft, A., Reed, C., Bourget, D., Lawrence, J., Börner, K., & Light, R. (2014). From big data to argument analysis and automated extraction: A selective study of argument in the philosophy of animal psychology from the volumes of the hathi trust collection. *Final Report. Digging into Data*.
- McNeill, K. L. (2009). Teachers' use of curriculum to support students in writing scientific arguments to explain phenomena. *Science Education*, 93(2), 233-268.
- McNeill, K. L., & Krajcik, J. (2007). Middle school students' use of appropriate and inappropriate evidence in writing scientific explanations. *Thinking with Data*, 233-265.
- Ministry of Education (MOE) (2015). 2015 revised science curriculum. Ministry of Education 2015-74 [issue 9].
- Nielsen, J. A. (2013). Dialectical features of students' argumentation: A critical review of argumentation studies in science education. *Research in Science Education*, 43(1), 371-393.
- Oh, J. A., Lee, S. -K., & Kim, C. -J. (2008). A case study on scientific inquiry and argumentative communication in earth science MBL classes. *Journal of the Korean Earth Science Society*, 29(2), 189-203.
- Ok, J. -S. (2004). A study on the prospect of using speech recognition softwares in elementary English education. *STEM Journal*, 5(2), 65-82.
- Osborne, J., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, 41(10), 994-1020.
- Paik, S. H., & Son, S. H. (2014). An analysis of pre-service science teachers' argument structures, the factors affecting the practice of state change experiment in 7th grade and cognition of pre-service education. *Journal of the Korean Association for Science Education*, 34(3), 197-206.
- Pardo, A. (2014). Designing learning analytics experiences. In J. A. Larusson & B. White (Eds.), *Learning analytics: From research to practice* (pp. 15-38). NY: Springer.

- Park, E. L., & Cho, S. (2014). KoNLPy: Korean natural language processing in Python. Proceedings of the 26th Annual Conference on Human & Cognitive Language Technology, Chuncheon, Korea, Oct 2014.
- Park, J. H., & Shin, N. M. (2017). Student's perceptions of artificial intelligence technology and artificial intelligence teachers. *The Journal of Korean Teacher Education*, 34(2), 169-192
- Park, S. -K. (2015). Analysis on the argumentation pattern and level of students' mental models in modeling-based learning about geologic structures. *Journal of the Korean Association for Science Education*, 35(5), 919-929.
- Ripley, B. (2016). tree: Classification and regression trees. R package version 1.0-37. Retrieved January 3, 2018, from <https://CRAN.R-project.org/package=tree>
- Rosé, C., Wang, Y. -C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., & Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *Computer-Supported Collaborative Learning*, 3, 237-271.
- Sampson, V., & Blanchard, M. R. (2012). Science teachers and scientific argumentation: Trends in views and practice. *Journal of Research in Science Teaching*, 49(9), 1122-1148.
- Sampson, V., & Clark, D. B. (2008). Assessment of the ways students generate arguments in science education: Current perspectives and recommendations for future directions. *Science Education*, 92(3), 449-472.
- Sampson, V., & Schleigh, S. (2013). *Scientific argumentation in biology: 30 classroom activities*. Arlington, VA: NSTA Press.
- Scheuer, O., McLaren, B. M., Loll, F., & Pinkwart, N. (2012). Automated analysis and feedback techniques to support and teach argumentation: A survey. In N. Pinkwart & B. M. McLaren (Eds.), *Educational technologies for teaching argumentation skills* (pp. 71-124). Bentham Books.
- Shin, H. S., & Kim, H. -J. (2011). Development of the analytic framework for dialogic argumentation using the TAP and a diagram in the context of learning the circular motion. *Journal of the Korean Association for Science Education*, 32(5), 1007-1026.
- Shin, S., Ha, M., & Lee, J. -K. (2017). High school students' perception of artificial intelligence: Focusing on conceptual understanding, emotion and risk perception. *Journal of Learner-Centered Curriculum and Instruction*, 17(21), 289-312.
- Simon, S., Erduran, S., & Osborne, J. (2006). Learning to teach argumentation: Research and development in the science classroom. *International Journal of Science Education*, 28(2-3), 235-260.
- Sionti, M., Ai, H., Rosé, C. P., & Resnick, L. (2011). A framework for analyzing development of argumentation through classroom discussions. *Educational technologies for teaching argumentation skills*, 28-55.
- Soller, A. (2004). Computational modeling and analysis of knowledge sharing in collaborative distance learning. *User Modeling and User-Adapted Interaction*, 14(4), 351-381.
- Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- Venables, W. N. & Ripley, B. D. (2002) *Modern applied statistics with S* (4th ed.). NY: Springer.
- Wee, S., Yoon, J., & Lim, S. (2014). An analysis on argumentation structure development of preservice teachers through argumentative writing on earth science related SSI. *Journal of the Korean Society of Earth Science Education*, 7(1), 11-23.
- Won, H. -Y. (2011). A study on the colloquial final ending '-잖아(요)' for Korean language education. *Korean Language & Literature*, 79, 307-327.
- Williams, J. M., & Colomb, G. G. (2007). *The craft of argument*. Pearson Education.
- Yu, Y. -L., & Baek, S. -G. (2016). Issue analysis of the related mass media's news articles on the 2015 revised national curriculum using automated text analysis. *The Journal of Curriculum and Evaluation*, 19(3), 127-156.
- Zhang, B. -T. (2017). *Deep learning*. Seoul: Hongrung Publishing Company.
- Zhao, M. (2016). A study for the adverb postpositions of 'ey', 'eyse' in the Korean language: Focused 'ey', 'eyse'. *Studies in Linguistics*, 40(7), 337-359.

[참고 자료 1] 머신 러닝을 위한 학습 데이터를 추출한 기존 문헌 목록 및 각 문헌의 논변 구조 분석 방법

문헌	논변 구조 분석 방법
1. Toulmin의 틀에 제시된 논변의 구성 요소를 그대로 따른 경우	
Lee <i>et al.</i> (2015)	<ul style="list-style-type: none"> 한 학생의 발화가 끝날 때까지를 분석 단위로 삼았으며, 그 안에 논변의 구성 요소가 여러 개 나올 경우 중복 코딩함. 과학 교육 전문가 2인이 각자 코딩한 후, 일치하지 않는 부분은 합의를 통해 일치시킴.
Baek <i>et al.</i> (2014)	<ul style="list-style-type: none"> 1차로 Toulmin의 틀에 따라 분석한 뒤, ‘주장-근거-추론’의 틀(Sampson, & Blanchard, 2012)을 통해 2차 분석하여 적절한 추론이 이루어졌는지 분석함 연구 방법에 분석 단위는 제시되지 않았으나, 연구 결과에서 한 논변을 구성하는 요소 별로 코딩된 결과가 제시됨. 연구자 2인이 합의를 통해 코딩한 후, 합의를 이끌어낸 사례들 중 일부를 다른 과학 교육 전문가 2인에게 검증받음.
Lim <i>et al.</i> (2010)	<ul style="list-style-type: none"> 문장 단위로 분석하였으며, 그 안에 논변의 구성 요소가 여러 개 나올 경우 중복 코딩함. 초등교사 1인, 과학 교육 전공 교사 5인, 과학 교육 전문가 3인이 분석하였으며, 분석자 간 일치도(inter-rater reliability)를 확인함.
Shin, & Kim (2011)	<ul style="list-style-type: none"> 학생들의 진술을 시간의 흐름에 따라 나열한 뒤, 논의과정에 해당하는 진술들만으로 자료를 축소함: 기존 문헌을 바탕으로 Toulmin이 제시한 구성 요소의 의미를 수정하여 코딩함. 여러 차례에 걸쳐 코딩하였으며, 전문가 1명과 과학 교육자 5명이 논의하며 분석이 이루어짐.
Lee (2006)	<ul style="list-style-type: none"> 한 학생의 발화가 끝날 때까지를 분석 단위로 삼았으며, 그 안에 논변의 구성 요소가 여러 개 나올 경우 중복 코딩함. 과학 교육 전문가 1인과 대학원생 1인이 각자 코딩한 후, 일치하지 않는 부분은 합의를 통해 일치시킴.
Wee <i>et al.</i> (2014)	<ul style="list-style-type: none"> 글쓰기 논변 활동에서 학생들이 작성한 글을 분석함. 문장 단위로 분석되었으며, 일련의 문장들이 하나의 요소에 해당하는 경우에는 하나의 요소로 분석함.
Lee <i>et al.</i> (2005)	<ul style="list-style-type: none"> 논변의 구성 요소가 드러나는 절 단위로 분석되었음.
Oh <i>et al.</i> (2008)	<ul style="list-style-type: none"> 한 학생의 발화가 끝날 때까지를 분석 단위로 삼았으며, 그 안에 논변의 구성 요소가 여러 개 나올 경우 중복 코딩함. 여러 차례에 걸쳐 코딩하였으며, 연구자 1인이 코딩한 뒤 그 적절성을 과학 교육 전문가 2인이 검토하고, 논의를 통해 의견이 일치되지 않는 부분은 논의를 통해 합의에 이를 때까지 수정함.
2. Toulmin의 틀에 제시된 논변의 구성 요소를 축소시킨 경우	
Paik, & Son (2014)	<ul style="list-style-type: none"> 예비교사들의 실험 보고서에서 드러나는 논증 구조를 분석함. 분석 자료에서 드러나지 않은 요소인 지지와 반박을 분석 틀에서 제외하고, 한정어의 의미를 ‘주장이 나타나지 않은 이유 혹은 요인’으로 확대함. 각 논변의 구성 요소에 해당하는 내용을 연구자들이 정리하여 제시함. 과학 교육 전문가 1인과 현장 교사 1인이 함께 분석하였으며, 이후 현장 교사 9인이 분석 내용을 검토함.
Lee <i>et al.</i> (2015)	<ul style="list-style-type: none"> Toulmin의 틀을 수정한 ‘주장-증거-추론’의 틀(McNeill, & Krajcik, 2007)을 활용하였으며, 이때 증거는 주장을 뒷받침하는 정보나 데이터를 의미하며, Toulmin의 틀에서 자료와 유사하며, 추론은 주장을 지지하기 위한 증거로서 데이터가 왜 중요한지 보여주는 정당화를 의미하며, Toulmin의 틀에서 보장과 유사함. 또한 이 틀에 반론과 반박을 추가함. 한 학생의 발화가 끝날 때까지를 분석 단위로 삼음. 연구자 1인과 과학 교육 전문가 1인이 코딩한 후, 일치하지 않는 부분은 합의를 통해 일치시킴.
Kwon, & Kim (2016)	<ul style="list-style-type: none"> Toulmin의 틀이 지닌 한계에 관한 선행 연구의 논의(Sampson, & Clark, 2008; Jiménez-Aleixandre <i>et al.</i>, 2000)에 따라 이 틀을 변형한 McNeill, & Krajcik(2007)의 틀을 참조하여 논변의 구성 요소를 주장, 자료, 정당화, 반박으로 제시함. 이때 정당화는 주장과 자료 사이의 관련성에 대한 진술이라고 정의함. 연구 방법에 분석 단위는 제시되지 않았으나, 연구 결과에서 한 논변을 구성하는 요소 별로 코딩된 결과가 제시됨. 과학 교육 전문가 2인이 각자 코딩한 후, 일치하지 않는 부분은 합의를 통해 일치시킴.
Park (2015)	<ul style="list-style-type: none"> Toulmin의 틀에서 자료, 주장, 보장, 보강, 반박을 사용하였으며, 대안개념이 포함되는 요소를 ‘대안개념의 자료’, ‘대안개념의 주장’과 같이 별도의 요소로서 구분하여 코딩함. 한 학생의 발화가 끝날 때까지를 분석 단위로 삼았으며, 그 안에 논변의 구성 요소가 여러 개 나올 경우 중복 코딩함. 연구자와 지도교사 2인이 각자 코딩한 후, 일치하지 않는 부분은 합의를 통해 일치시킴.
Lee <i>et al.</i> (2016)	<ul style="list-style-type: none"> Toulmin의 틀과 Sampson, & Clark(2008)의 분석틀을 참고하여 증거, 추론, 설명모델로 모델링에서의 분석틀을 고안함. 이때 증거는 Toulmin의 틀에서 자료에 해당하며, 추론은 보장, 지지, 반증, 한정어를 포함하는 요소이며, 설명 모델은 주장에 해당함. 연구 방법에 분석 단위는 제시되지 않았으나, 연구 결과에서 모델을 구성해나가는 추론 과정을 구성하는 요소 별로 코딩된 결과가 제시됨. 연구자 1인이 코딩한 후, 공동연구자들과 함께 검토하여 의견이 일치하지 않는 부분은 합의를 통해 일치시킴.
Cho <i>et al.</i> (2013)	<ul style="list-style-type: none"> Toulmin의 틀에서 자료, 주장, 보장을 사용하여 코딩함. 연구 방법에 분석 단위는 제시되지 않았으나, 연구 결과에서 모델을 구성해나가는 추론 과정을 구성하는 요소 별로 코딩된 결과가 제시됨. 연구자와 과학 교육 전문가 1인이 각자 코딩한 후, 일치하지 않는 부분은 합의를 통해 일치시킴.
Han <i>et al.</i> (2012)	<ul style="list-style-type: none"> Toulmin의 틀에 따라 분석하였으나, 본 연구의 목적에 따라 어떤 요소에 대하여 반박이 이루어졌는지 탐색하기 위해 연구 결과에서는 주장, 자료, 보장, 반박으로 코딩된 결과를 제시함. 한 학생의 발화가 끝날 때까지를 분석 단위로 삼았으며, 그 안에 논변의 구성 요소가 여러 개 나올 경우 중복 코딩함. 연구자 4인이 각자 코딩한 후, 일치하지 않는 부분은 합의를 통해 일치시킴.
Lee <i>et al.</i> (2012)	<ul style="list-style-type: none"> Toulmin의 틀에서 자료, 주장, 보장을 사용하여 코딩함. 연구 방법에 분석 단위는 제시되지 않았으나, 연구 결과에서 문장을 단위로 하여 한 논변을 구성하는 요소 별로 코딩된 결과가 제시됨.
Lee, & Kim (2011)	<ul style="list-style-type: none"> Toulmin의 틀에서 주장, 데이터, 보장을 사용하여 코딩함. 연구 방법에 분석 단위는 제시되지 않았으나, 연구 결과에서 한 학생의 발화가 끝날 때까지를 분석 단위로 삼아 코딩된 결과가 제시됨.

Lim, & Yeo (2012)

- 글쓰기 논변 활동에서 학생들이 작성한 글을 분석함.
 - Toulmin의 틀이 지닌 한계에 관한 선행 연구의 논의(Booth *et al.*, 2008; Williams, & Colomb, 2007)에 따라 이 틀을 변형한 Williams, & Colomb (2007)의 틀을 참조하여 논변의 구성 요소를 주장, 이유, 근거, 반론 수용과 반박으로 구분함. 이때 이유는 보증 이유와 주장을 이어주는 원칙이라고 정의하였으며, 근거는 Toulmin의 틀에서 자료에 해당함. 반론 수용과 반박은 주장에 대한 의문, 반대, 대안 또는 대안에 해당함.
 - 연구 방법에 분석 단위는 제시되지 않았으나, 연구 결과에서 각 요소별로 코딩된 결과가 제시됨.
 - 활동지를 무작위로 선정하여 과학 교육 전공자 6인과 함께 일치도를 확인하였고, 일치도가 0.9 이상이 된 이후에 연구자가 전체를 분석함.
-