인공지능 기반의 기초 데이터 과학 교육에 관한 연구

조정희
부산교육대학교 컴퓨터교육과

요약

데이터 과학은 문제를 정의하고 수집된 데이터로부터 의미 있는 정보를 찾아내어 문제를 해결하는 과학적인 학문이다. 인공지능 기술의 도입으로 인해 데이터의 활용 분야는 점차 확장되고 있으며, 데이터 과학 교육의 중요성에 대한 인식도 증가하고 있다. 그러나, 한국 데이터 산업진흥원에서 2020년 데이터 산업 현황을 조사하여 발표한 결과에 의하면 국내 데이터 산업의 시장규모가 급성장하고 있음에 불구하고, 향후 5년 이내에 데이터 전문가의 부족률은 31.4%에 이를 것으로 예측하였다. 이러한 사회적 분위기 속에서, 초등 교육 현장에서는 학생들의 컴퓨팅 사고력과 창의력 향상을 위하여 데이터 과학 교육을 도입하고자 다양한 연구가 진행되고 있다. 본 논문에서는 컴퓨터 분야의 비전공자가 대다수인 현직 초등교사들의 교육을 목적으로 개발하여 현장에서 활용하고 있는 데이터 과학 강의의 콘텐츠를 제안하였다. 해당 콘텐츠는 인공지능융합대학원에 재학 중인 현직 초등교사 집단을 대상으로 적용한 후, 학습자들이 이해하기에 어려움이 있었던 내용을 세부적으로 파악하고 원인을 분석하여 개선점을 도출하였다.

키워드 : 인공지능, 소프트웨어 교육, 데이터 과학, 초등 교육, 교육과정

A Study on Development of Basic Data Science Education Contents for Artificial Intelligence Capability

Junghee Jo
Dept. of Computer Education, Busan National University of Education

Abstract

Data science is a scientific discipline that defines problems while finding meaningful information from collected data to solve problems. Along with artificial intelligence technology, the field of data utilization is gradually expanding, and awareness of the importance of data science education is also increasing. Despite the rapid growth of the domestic data industry market, it has recently been predicted that the shortfall of data experts will reach 31.4% within the next 5 years according to an analysis of the current status of the data industry by the Korea Data Agency. In the field of elementary education, various studies have been conducted to introduce data science in order to improve students' computational thinking and creativity. This paper proposed the contents of data science lectures developed for the purpose of educating elementary school teachers, who are mostly non-majors in the computer field. The developed contents were applied to a group of elementary school teachers attending graduate school for artificial intelligence convergence education. Points for improvement were derived by identifying the contents that were difficult for learners to understand and analyzing the causes of difficulty.

Keywords : Artificial Intelligence, Software Education, Data Science, Elementary Education, Education Curriculum

논문투고 : 2021-08-04
논문심사 : 2021-08-04
심사완료 : 2021-08-06

1. 서론

인공지능 기술의 확산으로 인해 데이터의 활용 분야는 산업계 전반에 걸쳐 급속도로 확장되고 있다. 한국데이터 산업진흥원의 '2020 데이터 산업 현황 조사'의 연구보고서[1]에 의하면, 국내 데이터 산업의 시장규모는 2019년에 16조 8,582억 원으로 전년 대비 8.3% 성장했으며, 2020년은 19조 2,736억 원으로 추정하였다. 2020년 기준으로 데이터 산업에 종사하는 인력은 총 366,021명으로 2019년 대비 6.2% 증가했으며, 이 중 데이터 직무 인력은 101,967명으로 2019년 대비 14.5% 증가하였다. 향후 5년 이내에 기업에서 현재 인력보다 더 필요로 하는 인력은 12,114명으로 부족률이 10.6%로 파악되었으며, 고도의 전문성을 요하는 '데이터 과학자'의 부족률은 31.4%로 가장 높은 것으로 조사되었다. 데이터 과학자는 관련 업무를 수행하기 위해 기본적으로 컴퓨터 사이언스와 통계학 기술뿐만 아니라 분석하고자 하는 분야에 관한 전문지식을 충분히 갖추었다고 가정한다. 노규성(2015)은 데이터 과학자의 직무를 DACUM (Developing A Curriculum) 기법을 기반으로 분석하여 20가지의 지식과 15가지의 기술이 필요함을 증명하였고[2], 송현민(2015)은 데이터 과학자의 업무를 '컴퓨터 공학, 데이터 공학, 경영학, 통계학적인 지식과 기술을 활용, 데이터 기반의 정책 제도, 기획, 처리, 분석, 시각화, 운영관리 업무를 수행하고 빅데이터를 수집, 저장, 분석, 표현하는 체계를 만드는 지식과 기술을 습득하여 프로세스 혁신 및 신제품 개발, 마케팅 전략 결정 등의 과학적 의사결정을 끌어내는 직무'라고 정의하였다[3]. 즉, 데이터 과학자는 수집된 데이터에 관한 과학적인 관찰, 시스템적 분석, 가설 설정 및 검증을 수행하는 등의 고난도 작업이 가능해야 한다. 이에, 데이터 과학 교육의 중요성에 대한 인식이 확산하면서 관련 연구들도 급속히 증가하고 있다.

이러한 사회적 흐름에 맞추어, 초등 교육 현장에서도 학생들의 컴퓨팅 사고력과 창의성 향상을 목적으로 데이터 과학 교육을 도입하고자 다양한 연구가 진행되고 있다. 구덕회(2020)는 데이터 과학 교과 내용의 특성을 반영한 교육 프로그램을 개발하였다. 이 프로그램의 교수 학습 단계는 문제정의, 데이터 수집, 데이터 분석, 데이터 표현, 평가 및 일반화로 설계하였으며, 이를 적용하기 위해 차시별 세부 지도계획을 수립하고 예시 교재를 개발하였다[4]. 이와 유사한 연구사례로, 홍지연(2020)은 초등학생의 데이터 과학 교육 프로그램을 개발하여 이 프로그램이 학생들의 데이터에 대한 이해, 수집, 분석, 표현 능력의 향상에 도움이 되는지를 통계학적 방법을 통해 검증하였다[5]. 특정 도구를 활용한 데이터 과학 교육에 관한 연구는, 김용민(2020)은 앱 인벤터를 활용한 PBL(Project Based Learning) 기반의 데이터 과학 교육 수업을 설계하고 초등학생의 컴퓨팅 사고력과 창의성에 미치는 영향을 분석하였다. 이를 위해, ADDIE(Analysis, Design, Development,Implementation, Evaluation) 모형의 절차에 따라 PBL 기반의 데이터 과학 교육 수업을 설계하고, 초등학생들을 대상으로 42차시의 수업을 실시하여 통계학적 방법으로 컴퓨팅 사고력이 향상되었음을 검증하였다[6]. 허경(2020)은 블록 프로그래밍 기반의 엔트리에서 제공하는 데이터 수집, 가공, 및 분석 도구를 사용한 데이터 과학 교육을 제안하였다[7]. 이에, 본 논문에서는 현직 초등교사를 대상으로 개발하여 교육 현장에서 활용하고 있는 데이터 과학 수업의 구성을 설명하고, 수업에 참여하였던 학생들이 선호하는 수업의 내용과 향후 수업에의 적용 방안들을 분석하고 결과를 해석하였다.

2. 이론적 배경

2.1. 데이터 과학

데이터란, 관찰이나 측정 등을 통해서 얻은 단순한 사실이나 값을 의미한다[8]. 최근 들어, 사물 인터넷 기술의 빠른 발달로 다양한 종류의 센서가 발명되었고 이러한 센서로부터 많은 양의 데이터가 빠른 속도로 수집되고 있다. 또한, 뇌파와 같은 생체 데이터처럼 예전에는 수집이 쉽지 않았던 데이터까지 지금은 손쉽게 수집할 수 있어서 가용한 데이터의 종류와 양이 폭발적으로 증가함에 따라 '빅데이터'라는 용어도 등장하게 되었다. 한 예로, 노르웨이의 연구기관인 SINTEF는 현재 시점에서 수집된 데이터의 90%가 지난 2년간 만들어진 최신 데이터임을 밝혔다[9]. 과거 아날로그 환경에서 수집되는 데이터와 비교했을 때, 최근에 수집되는 데이터는 숫자 및 텍스트 데이터와 같은 정형 데이터를 비롯하여

이미지 및 비디오 데이터와 같은 비정형 데이터도 포함하므로 데이터의 종류와 양이 더욱 방대해진 것이다. 데이터 과학이란, 데이터에서 의미 있는 지식을 추출하는 학문이다. 즉, 지식을 얻고자 하는 문제를 정의하고, 문제를 해결하기 위한 데이터의 수집, 처리, 분석 방법을 설계하며, 분석된 결과물을 기반으로 지식을 추출하는 모든 과정을 포함한다. 데이터 과학은 통계학, 컴퓨터 과학, 정보기술, 도메인 특화 영역 등 여러 분야의 융합 학문으로써[10-12], 기존의 분석 방법으로는 처리가 어려운 데이터를 다양한 학문에서 제공하는 복합적인 기술을 적용하여 데이터를 단순히 분류하는 단계에서 진보하여 데이터 속에 담긴 패턴이나 예측에 도움이 되는 결과도 도출한다.

2.2. 데이터 과학 연구 동향

최근, 포브스는 2020년을 선도할 주요 기술 7가지를 발표하였는데 그중의 하나가 빅데이터이다[13]. 빅데이터는 RDBMS(Relational DataBase Management System)에 저장되지 않는 비정형의 데이터까지도 포함하는 대용량의 데이터로써 이러한 정보를 처리하기 위해서는 고도의 분석 기술이 필요하다. 예를 들어, 빅데이터의 많은 부분을 차지하고 있는 공간 빅데이터의 경우는 빅데이터의 일반적인 특징과 더불어 공간정보의 특성도 보유하므로 다른 종류의 빅데이터에 비하여 다양성과 복잡도가 매우 높은 특징을 가지고 있다[14]. 이러한 공간 빅데이터를 분석하기 위한 다양한 시스템이 활발히 연구되고 있는데 대표적인 것들로서 아파치의 하둡(Apache Hadoop), 아파치의 스파크(Apache Spark), 미국의 미네소타 대학에서 개발한 공간 하둡(SpatialHadoop)등이 있다. 하둡은 다수의 컴퓨터를 하나의 컴퓨터인 것처럼 묶는 기술을 통해 저장 공간과 계산 능력을 향상하게 시킴으로써 대용량의 데이터 처리를 가능하게 하는 자바 기반의 공개 소프트웨어 프레임워크이다. 이는 빅데이터를 분산 저장 관리하는 분산 파일시스템(HDFS, hadoop distributed file system)과 분석을 담당하는 분산처리시스템인 맵리듀스(MapReduce)로 구성되어 동작하며, 빅데이터를 처리하는 대표적인 프레임워크로 알려져 있다[15]. 스파크는 하둡의 한계를 보완하기 위한 인메모리 기반의 분산 데이터 분석 시스템으로써 데이터를 디스크에 쓰는 하둡과는 달리 메모리에서 처리하므로 속도가 빠른 장점이 있다[16]. 공간 하둡은 공간 빅데이터를 분석하기 위한 목적으로 기존의 하둡에 공간 정보 처리 기능을 추가하여, 공간 분석을 위해 많이 사용하는 영역 질의, 최단 경로 및 공간 조인과 같은 공간 분석 기능을 수행할 수 있도록 하였다[17].

2.3. 데이터 과학 교육 동향

데이터 과학이 현재 초, 중, 고등학교의 정보 교과에 적용된 최근 실태를 조사한 결과에 의하면, 정부에서 승인한 정보 교과서 중에 금성출판사의 경우에는 데이터와 가까운 주제인 '자료와 정보'가 전체 정보 교과 내용의 15%를 차지하고 있으며, 천재 교과서의 정보 교과서의 경우에는 20%를, 교학사의 정보 교과서의 경우에는 17%를 차지하고 있다. 이들 교과서는 공통으로 프로그래밍에 가장 큰 비중(31%-45%)을 두고 구성되어 있으며, 데이터와 관련해서는 실질적인 예시와 구체화한 내용보다는 간단한 이론적인 내용으로 대략적인 설명으로 교육을 하는 것으로 조사되었다[18]. 데이터 과학의 교육과 관련된 최근 연구에 의하면, 이명호(2016)는 국내외에 증가하고 있는 데이터 과학 프로그램에 대한 교과과정을 조사하고, 국내와 국외의 데이터 과학 교과과정과 비교하여 국내 교과과정의 개선 방향을 제시하였다[19]. 이와 유사한 맥락으로, 이혜원(2020)은 2019년 10월을 기준으로 데이터 과학과 관련하여 국내에 개설된 80개의 교육과정과 2,041개의 교과목을 대상으로 학문 영역 특징 기반 분석, 데이터 전문가 역량 기반 분석, 그리고 교과목명 내용 분석을 시행하였다[20]. 김재경(2021)은 컴퓨터 분야의 비전공자를 대상으로 컴퓨팅 사고력을 향상하는 목적으로 데이터 과학을 교양 과목으로 편성하고 ADDIE 모델을 기준으로 수업안을 개발하였다. 그리고 대학생을 대상으로 15주간 교육하고 사전 사후 자기 평가 설문을 시행하여 결과를 분석하고 교육의 효과가 있음을 입증하였다[21].

3. 데이터 과학 수업 설계

3.1. 이론적 측면

본 연구를 통해 개발된 데이터 과학 수업 콘텐츠는 현직 초등교사를 대상으로 하였다. 이론의 초반에는 '데이터'의 개념에 관해 설명하였으며, 특히 빅데이터의 생성 배경을 사물 인터넷 및 인공지능과 연결 지어 설명하였으며 이 과정에서 크게 이바지한 대표적인 과학자들에 대해 설명하였다. 예를 들어, 많은 양의 이미지 데이터베이스인 이미지넷을 구축한 스탠퍼드대학의 페이페이 리 교수의 업적을 인공지능과 데이터 과학을 연결지어 설명함으로써 학습자들이 두 기술 간의 연관성을 이해하도록 유도하였다. 이어서, '데이터 과학'이라는 학문의 정의 및 배경지식, 그리고 요소 기술들을 소개하였고, EBS에서 제작한 데이터 과학의 개념을 설명하는 짧은 동영상으로 설명을 마무리하였다.

다음으로, '데이터 과학자'의 정의를 설명하고 학습자들이 익숙한 '나이팅게일'의 예를 들어 데이터 과학자가 주로 수행하는 업무들을 설명하였다. 그리고, 국내외의 현장에서 데이터 과학자로 일하고 있는 다양한 사례들을 소개해줌으로써 현실감을 갖도록 하였다. 예를 들어, 국내의 경우는 특정 대기업의 인재 영입 사이트에 접속해서 데이터 분석과 관련된 다양한 인력 모집 공고들을 학습자들에게 보여줌으로써 산업계에서 필요로 하는 데이터 과학자의 급진적인 수요에 대해 직접 느끼도록 하였다. 국외의 경우는 컴퓨터 분야의 선진적인 기업에서 현재 데이터 과학자로 근무하고 있는 한국인의 인터뷰 동영상을 통해 다양한 경험을 간접적으로나마 체험할 기회를 제공하였다.

이후에는 데이터 과학의 절차에 대한 상세한 설명이 이어지고 마지막으로 교수자가 직접 겪은 다양한 데이터의 수집 및 분석 프로젝트들의 내용을 학습자들과 공유하고 이를 기반으로 토론하는 시간을 가짐으로써 이론으로 배운 데이터 과학과 실제로 현업에서 이루어지는 데이터 과학의 내용을 스스로 생각해보고 연결하도록 하였다.

3.2. 실습적 측면

빅데이터를 대상으로 R, 파이썬, 및 데이터베이스를 활용한 실습수업을 위해서는 기본적으로 텍스트 프로그래밍 언어와 데이터 분석을 위한 통계학의 이해가 수반된다. 엔트리는 블록 코딩 방식의 소프트웨어 교육 플랫폼으로 컴퓨터 분야의 비전공자인 초등교사에게도 친숙한 도구이다. 엔트리는 개발된 이후 꾸준히 업그레이드되고 있으며 2020년 2월에는 데이터를 수집하고, 시각화하고, 프로그래밍과 접목하여 활용하는 기능이 추가됨으로 인해 초등 교육 현장에서도 데이터 과학의 교육이 가능하게 되었다.

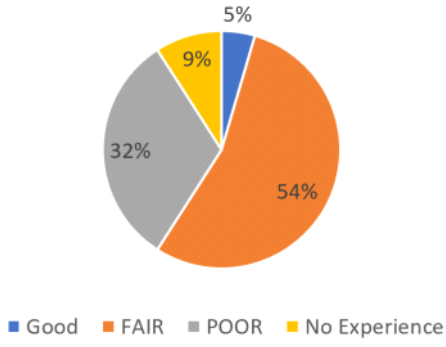
본 연구를 통해 개발된 수업 콘텐츠의 전반부에는 앞에서 설명한 데이터 과학의 이론 학습으로 구성하였고, 후반부에는 엔트리의 데이터 과학 기능을 활용한 실습으로 구성하였다. 엔트리 데이터 분석을 활용하기 위해서는 기본적으로 데이터가 '테이블'의 형태로 제공되어야 하므로, 실습의 초반에 테이블의 개념을 설명하였고 '행', '열', '레코드', '속성' 등의 테이블과 관련된 용어에 관해 설명을 하였다. 데이터 준비 방법은 3가지로써, '새로 만들기' , '테이블 선택', '파일 올리기'가 있는데 난이도의 순서대로 예시와 함께 설명하였다. 먼저, '직접 만들기'의 경우는 사용자가 엔트리에 직접 빈 테이블을 생성하여 행과 열을 추가하고 데이터를 입력하는 방법인데 (Fig. 1)의 '학교알리미' 사이트에서 제공하는 성별 학생 수에 관한 공공 데이터를 참조하여 엔트리에 데이터를 직접 입력하고 간단한 차트를 생성하여 시각화해 보도록 하였다.

(Fig. 1) Example of government provided open data

이어서, '테이블 선택'의 경우는 '네이버 블로그 평균 사용 시간', '연도별 배추 생산량' 등과 같이 엔트리에서 제공하는 테이블 형식의 데이터를 선택하고 이를 활용하여 초등수업에서 활용 가능한 간단한 엔트리 프로그램을 작성해보도록 하였다. 마지막으로, '파일 올리기' 기능을 활용하여 외부로부터 수집한 공공 데이터를 엔트리로 불러와서 활용하는 방법을 실습해보도록 하였다. 이후부터는, 다양한 종류의 데이터를 수집, 분석하고 변수 활용 및 신호 보내기와 같은 기능을 활용하여 여러 교과 수업에서 활용 가능한 엔트리 프로그램을 작성해 보도록 하였다.

4. 연구 결과

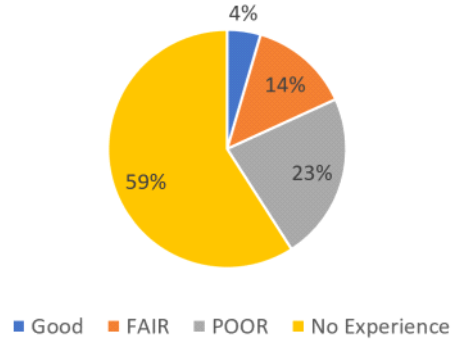
본 연구를 통해 개발된 데이터 과학 수업 콘텐츠는 인공지능융합대학원에 재학 중인 현직 초등교사 22명을 대상으로 15차시의 수업에 사용되었으며, 종강 후에 학습자들로부터 수집한 설문을 바탕으로 결과를 도출하였다.



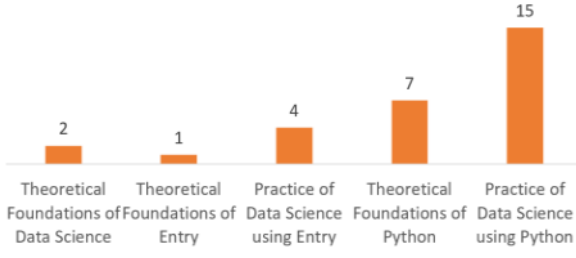
(Fig. 2) Analysis of block programming experience

(Fig. 2)는 학습자들이 대학원에 입학하기 전에 엔트리 또는 스크래치와 같은 블록 프로그래밍에 대한 경험치를 묻는 질문에 대한 응답 결과이다. 각 학습자는 다음의 4단계 중에 한 가지를 선택하도록 하였다. '상'은 원하는 프로그램 개발이 자유자재로 가능한 단계이고, '중'은 블록 프로그래밍의 이론적인 부분은 충분히 알고 있으나 프로그래밍은 미흡한 가능한 단계이며, '하'는 블록 프로그래밍의 이론적인 부분도 미숙하여 프로그래밍이 어려운 단계이다. 마지막으로, '없음'은 블록 프로그래밍을 접해보지 못한 단계를 의미한다. 가장 많은 응답을 얻은 단계는 '중'으로 총 22명 중에 54%의 학습자들이 본인이 이 단계에 해당한다고 응답하였다. 그 뒤로 '하' (32%), '없음' (9%), '상' (5%) 순으로 조사되었다.

(Fig. 3)은 학습자들이 대학원에 입학하기 전에 파이썬과 같은 텍스트 프로그래밍에 대한 경험치를 묻는 질문에 대한 응답 결과이다. 각 학습자는 앞서 설명한 4단계 중에 한 가지를 선택하도록 하였다. 가장 많은 응답을 얻은 단계는 '없음'으로 총 22명 중에 59%의 학습자들이 본인이 이 단계에 해당한다고 응답하였다. 그 뒤로 '하' (23%), '중' (14%), '상' (4%) 순으로 조사되었다. 특히, '상'이라고 응답한 1명의 학습자는 교육대학교에 입학하기 전, 공대의 기계공학전공 과정을 이수하면서 포트란, C++, 비주얼 베이직, 자바, SQL 등 다양한 종류의 텍스트 프로그래밍 언어를 접해본 경우였다.

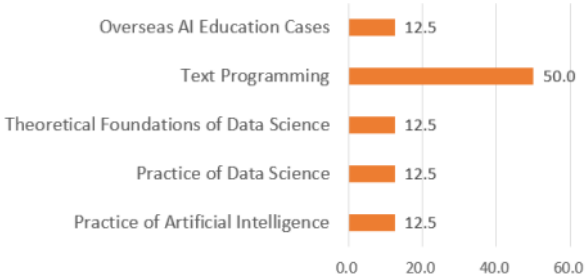


(Fig. 3) Analysis of text programming experience



(Fig. 4) Analysis of contents that learners had difficulty understanding

(Fig. 4)는 데이터 과학 강의 전체 콘텐츠 중에 학습자들이 이해하기에 어려움을 겪은 부분들에 관하여 묻는 질문에 대한 응답 결과이다. 이를 위해, 전체 강의 내용을 총 13가지의 내용으로 분류하였고, 학습자들은 복수개의 내용을 선택할 수 있도록 하였다. 13개 문항에 대한 학습자들의 선택 결과는 '데이터 과학 이론', '엔트리 이론', '엔트리 기반의 데이터 과학 실습', '파이썬 이론', '파이썬 기반의 데이터 과학 실습'의 5가지 카테고리로 합산하여 결과를 도출하였다. 가장 많은 응답을 얻은 내용은 '파이썬 기반의 데이터 과학 실습' (15명) 이었고, 이어서 '파이썬 이론' (7명), '엔트리 기반의 데이터 과학 실습' (4명), '데이터 과학 이론' (2명), '엔트리 이론' (1명) 순으로 조사되었다.



(Fig. 5) Categories of contents needing improvement to be addressed in future lectures

(Fig. 5)는 향후에 강의에서 보완이 필요한 부분을 묻는 질문에 대한 응답 결과이다. 총 22명의 학습자 중에서 구체적으로 응답한 8명(36%)을 대상으로 결과를 도출하였다. 가장 많은 수의 학습자들이 텍스트 프로그래밍 언어라고 응답하였으며 (50%), 그 외에 해외 인공지능 관련한 초등 교육 사례 (12.5%), 데이터 과학의 이론 (12.5%), 데이터 과학의 실습 (12.5%), 인공지능 교육 도구의 실습 (12.5%)와 같은 응답들이 수집되었다.

5. 결론

본 논문에서는 컴퓨터 분야의 비전공자가 대다수인 초등교사들을 대상으로 한 데이터 과학 수업 콘텐츠를 개발하였다. 그리고 해당 콘텐츠를 인공지능융합대학원에 재학 중인 현직 초등교사 집단을 대상으로 적용한 후 다음과 같은 결론을 도출하였다.

첫째, 현직 초등교사들은 교육대학교 재학 때 교과 수강 또는 교육 현장에서의 경험으로 인하여 블록 프로그래밍에 대한 경험이 풍부할 것이라고 예상하였으나, 실제로는 프로그래밍이 힘들거나 경험이 전혀 없는 학습자들이 상당수 존재하였다. 본 논문에서 제안한 수업 콘텐츠는 초등교사들이 중급 이상의 블록 프로그래밍 경험이 있을 것으로 가정하고 제작한 것이었으나, 향후에는 이를 보완하여 블록 프로그래밍의 개론에 관한 수업 시수를 추가하여야 할 필요가 있음을 파악하였다.

둘째, 데이터 과학의 실습을 위해 엔트리 이외에 텍스트 프로그래밍을 활용하는 방법을 전달하기 위해, 간단한 파이썬 문법을 활용한 시각화 생성에 관한 수업을 후반부의 2차시에 걸쳐서 실험적으로 실시하였다. 코로나로 인한 비대면 강의의 특성상, 학습자들의 이해도를 즉시 파악할 수 없었으므로 텍스트 프로그래밍에 대한 경험치를 묻는 설문을 시행하였고 예상대로 절반 이상의 학생들이 해당 경험이 전혀 없었다. 이는, 파이썬의 기초 이론과 시각화 실습이 단순한 내용임에도 불구하고 상당수의 학습자가 어려움을 느꼈던 원인으로 파악하였다.

마지막으로, 파이썬을 이용한 데이터 과학 실습에 어려움을 겪은 학습자들이 많음에도 불구하고, 향후에 보완이 필요한 강의 내용으로 상당수의 학습자가 '텍스트 프로그래밍'으로 응답하였다. 비록, 어려움은 느꼈으나 어느 정도 익숙한 엔트리보다는 새로운 분야인 텍스트 프로그래밍을 학습해보고자 하는 학생들이 상당수 존재함을 파악하였다.

참고문헌

[1] Korea Data Agency(2020). 2020 Data Industry Survey, Korea Data Agency.

[2] Noh, K.S., Park, S.T., Park, K.H.(2015). Convergence Study on Big Data Competency Reference Model. Journal of digital Convergence, 13(3), 55-63.

[3] Song, H.M.(2015). A Study of Development for job Competency in the bigdata using DACUM method, Korea University master's thesis.

[4] Koo, D.H., Kim D.J.(2020). Data science education program based on problem solving learning. Korean *Journal of Elementary Education, 31*, 203-215.

[5] Hong, J.H., Kim, Y.S.(2020) Development of AI Data Science Education Program to Foster Data Literacy of Elementary School Students, *Journal of The Korean Association of Information Education, 24*(6), 633-641.

[6] Kim, Y.M.(2020) The Effects of PBL-based Data Science Education classes using App Inventor on elementary student Computational Thinking and Creativity improvement, *Journal of The Korean Association of Information Education, 24*(6), 551-562.

[7] Hur, K.(2020). A Study on Elementary Education Examples for Data Science using Entry, *Journal of The Korean Association of Information Education, 24*(5), 473-481.

[8] Retrieved from https://en.wikipedia.org/wiki/Data.

[9] SINTEF(2013). Big Data, for better or worse: 90% of world's data generated over last two years. Retrieved from https://www.sciencedaily.com/releases/2013/05/130522085217.htm.

[10] Shi, Y., Philip, S.Y., Zhu, Y., Tian, Y.(2014). Explore New Field of Data Science Under Big Data Era: Preface for ICDS 2014. *Procedia Computer Science, (30)*, 1-3.

[11] Bruce, P., Bruce, A.(2017). Practical statistics for data scientists: 50 essential concepts. O'Reilly Media, Inc.

[12] O'Neil, C., Schutt, R.(2013). Doing data science: Straight talk from the frontline. O'Reilly Media, Inc.

[13] Forbes(2019). The 7 Biggest Technology Trends That Will Transform Telecoms In 2020. Retrieved from https://www.forbes.com.

[14] Jo, J.H., Lee, K.W.(2018). Marmot: A Hadoop-based High Performance Data Storage Management System for Processing Geospatial or Geo-Spatial Big Data. *Journal of Korean Society for Geospatial Information Science, 26*(1), 3-10.

[15] White, T.(2012). Hadoop: The Definitive Guide, 3rd ed.; O'Reilly Media, Inc. Sebastopol, CA, USA: 2012; ISBN 1449338771.

[16] Zaharia, M., Chowdhury, M., Franklin, M.J., Shenker, S., Stoica, I.(2010). Spark: Cluster Computing with Working Sets, *Proceedings of the 2nd USENIX conference on Hot topics in cloud computing, 10*, 10-10.

[17] Eldawy, A.(2014). SpatialHadoop: Towards flexible and scalable spatial processing using MapReduce. *Proceedings of the SIGMOD PhD symposium 2014, Snowbird, UT, USA, 22 June 2014*, 46 - 50.

[18] Park, H.S.(2020). A*nalyzing the current status of how education is conducted for data science, suggesting directionality, and present the data science integration class plan*. Inchon University master's thesis.

[19] Lee, M.H.(2016). A Study on the Curriculums of Data Science, *Journal of the Korean Biblia Society for Library and Information Science, 27*(1), 263-290.

[20] Lee, H.W., Han, S.H.(2020). An Analysis of Data Science Curriculum in Korea. *Journal of the Korean Society for Library and Information Science, 54*(1), 365-385.

[21] Kim, J.K, Sohn, E.S.(2021). Development of Data Science Course and Analysis of Computational Thinking Effect for non-Major Students. *The Journal of Korean Association of Computer Education, 24*(3), 23-31.

저자소개

조 정 희

2014년 University of Massachusetts Amherst, Computer Science, 공학박사

2005~2019 한국전자통신연구원(ETRI), 책임연구원

2019~현재 부산교육대학교 컴퓨터교육과 조교수

관심분야: 컴퓨터교육, 인간공학, 헬스케어

e-mail: dreamer@bnue.ac.kr