JPEE 실천공학교육논문지
JOURNAL OF PRACTICAL ENGINEERING EDUCATION

실천공학교육 관련
J. Pract. Eng. Educ. 15(1), 193-198, 2023

머신러닝 모델을 이용한 파이썬 자동채점 연습문제의 타당성 분석

Validity Analysis of Python Automatic Scoring Exercise-Problems using Machine Learning Models

허 경*

경인교육대학교 컴퓨터교육과

Kyeong Hur*

Department of Computer Education, Gyeong-In National University of Education, Anyang 13910, Korea

[요 약]

본 논문은 파이썬 프로그래밍 교육에서 단원별 연습문제의 타당성을 분석하였다. 단원별로 제시되는 연습문제는 온라인 학습 시스템을 통해 제시되고 학생 각자가 답안 코드를 업로드하여 자동으로 채점된다. 한학기 동안 진행되는 파이썬 교육을 통해, 학생들의 중간시험점수, 기말시험 점수 그리고 각 단원별 연습문제 점수 등 데이터가 수집된다. 수집된 데이터들을 통해, 자동채점 연습문제들의 타당도를 분석하여 단원별 연습문제들을 개선할 수 있다. 본 논문에서는 자동 채점 연습문제들의 타당도를 분석하기 위해, Orange 머신러닝 도구를 사용하였다. 파이썬 과목에서 수집된 데이터를 전체, 상위권 그리고 하위권 그룹별로 4가지 분석을 실시하고 종합적으로 비교한다. 파이썬 단원별 연습문제 점수들로부터 학생의 최종 성적을 예측하는 머신러닝 모델의 예측 정확도로부터 단원별 자동채점 연습문제의 출제 타당도를 분석하였다.

[Abstract]

This paper analyzed the validity of exercise problems for each unit in Python programming education. Practice questions presented for each unit are presented through an online learning system, and each student uploads an answer code and is automatically graded. Data such as students' mid-term exam scores, final exam scores, and practice questions scores for each unit are collected through Python lecture that lasts for one semester. Through the collected data, it is possible to improve the exercise problems for each unit by analyzing the validity of the automatic scoring exercise problems. In this paper, Orange machine learning tool was used to analyze the validity of automatic scoring exercises. The data collected in the Python subject are analyzed and compared comprehensively by total, top, and bottom groups. From the prediction accuracy of the machine learning model that predicts the student's final grade from the Python unit-by-unit practice problem scores, the validity of the automatic scoring exercises for each unit was analyzed.

Key Words: AI, Exercise problems, Machine learning, Model, Orange tool, Python programming

http://dx.doi.org/10.14702/JPEE.2023.193

Received 1 April 2023
Accepted 15 April 2023

***Corresponding Author**
E-mail: khur@ginue.ac.kr

I. 서 론

파이썬 프로그래밍 교육에 있어 학생들이 스스로 정답 코드를 탐색할 수 있는 환경을 구현하기 위해서는, 온라인으로 단원별 연습문제를 제공하고 자동 채점 및 피드백이 가능해야 한다. 또한, 다수 학생들의 연습문제 답안 평가를 자동화하여, 교수자의 부담을 줄이는 효과가 있다[1]. 채점 데이터는 입력 데이터와 정답 데이터로 구성되며, 자동 평가 프로그램은 학생이 업로드한 답안 코드를 실행하여 오류 여부를 확인한다. 학생의 코드에 오류가 있으면 코드 오류로 판정하고, 코드의 오류가 없으면 지정된 입력 데이터를 입력하여 실행하고 지정된 정답 데이터와 비교하여 정답 또는 오답을 판정한다. 문제 제시, 코드 작성, 피드백 등이 이루어지는 사용자 환경을 웹 사이트로 제공하고 있다[2]. 프로그래밍 자동 채점 시스템은 프로그래밍 교육이나 정보 올림피아드 등 경시대회에서 활발하게 사용되고 있다[3]. Orange는 오픈 소스를 기반으로 머신러닝 및 데이터 시각화 도구로서, 다양한 도구 상자를 사용하여 시각적으로 데이터 분석 워크플로를 구축하고 머신러닝 모델을 만들 수 있다[4].

자동 채점 시스템을 활용한 프로그래밍 교육에 대한 선행 연구를 살펴보면, 참고문헌 [3]에서는 자동 채점 시스템을 사용한 학생들이 교재의 소스 코드를 그대로 입력하며 학습한 학생들보다 프로그래밍 시험 점수가 높게 측정됨을 분석하였다. 참고문헌 [5]에서는 자료구조 강의를 위한 문제를 개발하고 자동 채점 시스템에 탑재하여 대학생에게 적용하였고 수업 태도, 코딩 흥미, 몰입과 경쟁에서 긍정적인 영향을 나타냄을 기술하였다.

참고문헌 [6]에서는 프로그래밍 자동 채점 시스템을 활용한 학생들이 전통적인 프로그래밍 수업을 실시한 학생들보다 정의적 영역에서 우수함을 보였고 프로그래밍 자동 채점 시스템이 몰입에 효과적이라는 것을 확인하였다 참고문헌 [7]에서는 자동 채점 시스템의 활용이 학습 동기와 컴퓨팅 사고력에 미치는 구조적 관계를 규명하였다. 학생의 학습목표 지향성과 자동 채점 시스템의 자동 채점 횟수는 자기 효능감과 몰입을 매개로 컴퓨팅 사고력 성취도에 효과를 나타내었다. 참고문헌 [8]에서는 비전공자를 위한 파이썬 기초 프로그래밍 커리큘럼과 파이썬 요소별 평가 문제 사례를 제안하고 유효성을 분석하였다.

이러한 선행 연구에 기초하여, 본 논문에서는 파이썬 프로그래밍 교육에서 단원별 연습문제의 타당성을 분석하였다. 단원별로 제시되는 연습문제는 온라인 학습 시스템을 통해 제시되고 학생 각자가 답안 코드를 업로드하여 자동으로 채점된다. 한학기 동안 진행되는 파이썬 교육을 통해, 학생들의 중간시험점수, 기말시험 점수 그리고 각 단원별 연습문제 점수 등 데이터가 수집된다. 수집된 데이터들을 통해, 자동채점 연습문제들의 타당도를 분석하여 단원별 연습문제들을 개선할 수 있다. 본 논문에서는 자동 채점 연습문제들의 타당도를 분석하기위해, Orange 머신러닝 도구를 사용하였다. 파이썬 과목에서 수집된 데이터를 전체, 상위권 그리고 하위권 그룹별로 4가지 분석을 실시하고 종합적으로 예측 시나리오를 분석한다. 파이썬 단원별 연습문제 점수들로부터 학생의 최종 성적을 예측하는 머신러닝 모델의 예측 정확도로부터 단원별 자동채점 연습문제의 출제 타당도를 분석하였다.

II. 파이썬 자동 채점 연습 문제와 머신러닝 모델 구축 워크플로우

A. 파이썬 자동 채점 연습 문제

참고문헌 [2]의 온라인 파이썬 학습시스템에서 지원하는 파이썬 자동 채점 연습 문제를 단원별로 분류하고 단원별 연습문제 수와 함께 표 1에 나타내었다. 2022년 파이썬 프로그래밍 과목에서는 단원1에서 단원4까지 진행한 후 중간시험을 실시하였고, 단원8까지 모두 강의한 후 기말시험을 실시하였다. 표 2에는 단원1에서 단원4까지 해당하는 자동채점 연습 문제들 중 하나씩 예시한 것이다. 단원별 연습문제는 기초적인 수준으로 작성된 것을 알 수 있다. 표 3에는 단원5에서 단원8까지 해당하는 자동 채점 연습 문제들 중 하나씩 예시한 것으로 기초 내용을 담고 있음을 알 수 있다.

표 1. 파이썬 자동 채점 연습 문제 출처 단원

Table 1. Python auto-scoring exercise problems source unit

단원	단원명	자동 채점 연습문제 수
단원1	• 출력, Print 단원	7
단원2	• 입력과 변수 단원	10
단원3	• 리스트와 문자열 단원	6
단원4	• 조건문 단원	5
단원5	• 반복문 for문 단원	5
단원6	• 반복문 while문 단원	6
단원7	• 함수 단원	4
단원8	• 모듈 단원	5

표 2. 단원1-단원4 자동 채점 연습 문제 예시

Table 2. Examples of unit 1-4 auto-scored practice questions

단원	문제내용	입력안내	출력안내	예시입력	예시출력
단원1	줄을 바꿔 문장을 출력해봅시다. 다음의 문장을 출력하세요. (대소문자에 주의해주세요) Hello Codle	입력 없음	Hello Codle을 출력합니다.	입력 없음	Hello Codle
단원2	3개의 정수(integer)를 입력받아 합과 평균을 각각 한 줄씩 출력하는 프로그램을 작성해봅시다.	정수 3개가 공백을 두고 입력됩니다.	세 수의 합 세 수의 평균	4 5 6	15 5.0
단원3	하나의 List를 입력받은 다음, 그 List의 최댓값과 최솟값의 합을 출력하는 프로그램을 작성하세요.	List는 띄어쓰기를 통해 분리된 값으로 입력됩니다.	List내의 숫자 중 최대값과 최소값의 합이 출력됩니다.	1 2 3 7 5 4	8
단원4	2개의 정수(integer) (a,b)를 입력받아 a와 b 가 다르면 True를, a 와 b가 같으면 False를 출력하는 프로그램을 작성해봅시다.	두 정수(a, b)가 공백을 두고 입력됩니다.	a와 b가 다른 경우 True 를, 그렇지 않은 경우 False 를 출력합니다.	0 1	True

표 3. 단원5-단원8 자동 채점 연습 문제 예시

Table 3. Examples of unit 5-8 auto-scored practice questions

단원	문제내용	입력안내	출력안내	예시입력	예시출력
단원5	1개의 정수(integer) (1~100)를 입력받아 거꾸로 카운트다운하는 프로그램을 작성해봅시다. (0부터 시작해 입력받은 정수까지 1씩 증가시키며 출력합니다.)	1~100의 정수 1개가 입력됩니다.	0부터 시작해 입력받은 정수까지 1씩 증가시키며 한 줄에 한개씩 출력합니다.	3	0 1 2 3
단원6	1개의 정수(integer) (0~1000)를 입력받아 입력받은 정수보다 작은동안 1, 2, 3 … 을 계속 더하는 프로그램을 작성해봅시다. 마지막에 더한 정수를 출력합니다.	정수 1개가 입력됩니다.	자연수를 순서대로 계속 더해 입력된 정수와 같거나 커졌을 때, 마지막에 더한 정수를 출력한다.	55	10
단원7	3개의 정수(integer)를 입력받아 합과 평균을 각각 한 줄씩 출력하는 프로그램을 작성해봅시다.	정수 3개가 공백을 두고 입력됩니다.	세 수의 합 세 수의 평균	4 5 6	15 5.0
단원8	2개의 정수(integer)를 입력받아 첫 번재 정수를 두 번째 정수번 거듭제곱한 값을 출력하는 프로그램을 작성해봅시다.	2개의 정수(n1, n2)가 공백으로 구분되어 입력됩니다.	n1을 n2번 거듭제곱한 값을 출력합니다.	2 10	1024

B. 머신러닝 모델 구축 워크플로우

표 4는 파이썬 단원별 연습문제 점수들로부터 학생의 최종 성적을 예측하는 머신러닝 모델을 구축하는 데 사용된 데이터 속성들을 설명하고 있다. 학습데이터 세트는 단원1점수부터 단원8점수까지 원인에 해당하는 설명변수 8개를 갖고, 결과에 해당하는 목적변수로 정규시험성적평균, 기말시험점수, 중간시험점수 그리고 최종석차 4개를 포함한다. 전체 12개 값으로 구성된 학습 데이터 수는 31개이며, 대부분 0~100 중 하나의 값을 갖고 최종석차는 1~31 중 하나의 값을 갖는다. 본 학습데이터는 22년에 실시한 파이썬 프로그래밍 과목에서 수집한 31명의 성적데이터이다.

표 4. 수치 데이터 머신러닝 모델 학습 및 최종성적 예측에 사용된 데이터 속성

Table 4. Data features used to train and predict numerical data machine learning models

학습데이터명	속성	데이터 수	수치값 범위
단원 1점수~단원 8점수	설명변수	31	0~100
정규시험성적평균	목적변수	31	0~100
기말시험점수	목적변수	31	0~100
중간시험점수	목적변수	31	0~100
최종석차	목적변수	31	0~31

학습 데이터 세트 중 단원1점수~단원8점수 8개값들로부터 목적변수, 즉, 정규시험성적평균, 기말시험점수, 중간시험점수 그리고 최종석차 4가지 값을 예측하는 머신러닝 모델을 만든다. 그림 1은 Orange로 작성한 수치 데이터 머신러닝 모델 학습 및 최종성적 예측 워크플로우를 나타낸다. Training Data 위젯을 통해, 12개 값을 갖는 31개 학습데이터가 입력된다. Data Table 위젯을 통해 입력된 학습데이터 세트를 확

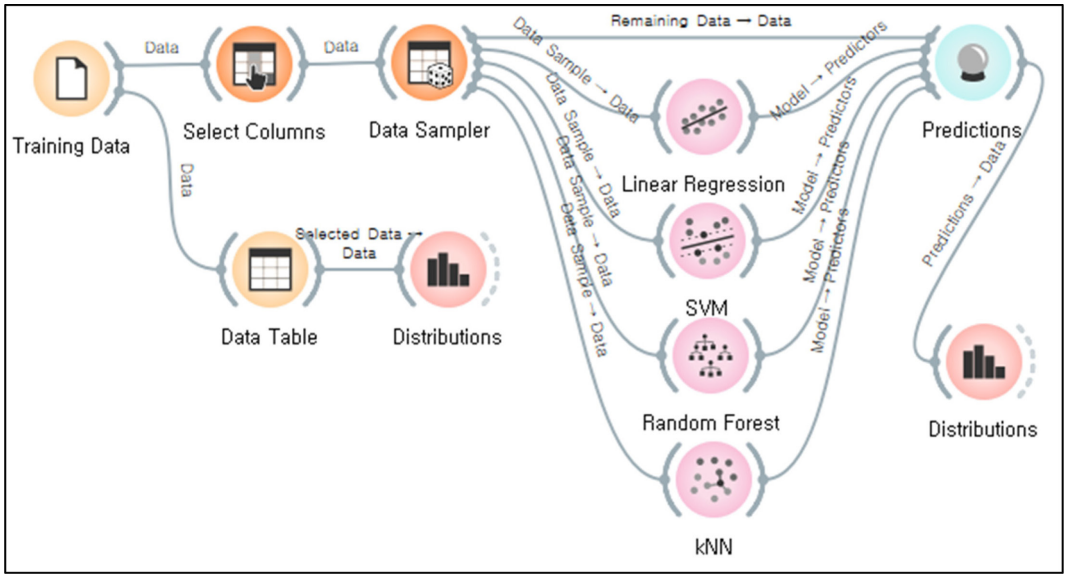

그림 1. 수치 데이터 머신러닝 모델 학습 및 최종성적 예측 워크플로우

Fig. 1. Numerical data machine learning model training and final grade prediction workflow.

인할 수 있고 Distribution 위젯을 통해, 위 12개 각 속성에 대한 도수분포 그래프를 확인할 수 있다. 학습데이터와 연결된 Select Columns 위젯에서는 입력된 학습데이터 중 다수 설명변수와 목적변수 1개를 지정한다. Data Sampler 위젯은 31개 학습데이터들 중 머신러닝 모델들에 입력되는 학습용 데이터와 테스트용 데이터의 비율을 결정하고, 실제 학습용 및 테스트용 데이터들을 추출한다. 그림 1에서 설정한 비율은 70%이며, 31개 학습데이터 중 실제 모델 학습에 사용된 데이터 수는 22개, 테스트용 데이터 수는 9개이다. 그림 1에서는 연속적인 수치 데이터들을 분석할 수 있는 머신러닝 모델로 Linear Regression, SVM, Random Forest와 KNN 모델들을 사용하였다.

III. 그룹별 모델 예측 정확도 분석

표 5는 온라인 자동 채점을 통해 수집된 31명 학생의 단원별 연습문제 점수 8개로부터 최종성적 수치데이터를 예측하는 시나리오 Case를 분류한 것이다. 즉, 그림 1에 나타낸 워크플로우로부터 C1~C4 4종류 머신러닝 모델들이 생성된다.

그림 1에서는 Linear Regression, SVM, Random Forest와 KNN 모델들을 사용하였다. 테스트 결과로부터 Linear Regression과 SVM 모델은 상대적으로 큰 RMSE(Root Mean Square Error) 값들을 나타내어, 최종적으로 Random Forest

표 5. 최종성적 예측 시나리오 Case 분류

Table 5. Final grade prediction scenario case classification

Case명	설명변수	목적변수
C1	단원1점수~단원8점수 8개	정규시험성적평균
C2	단원5점수~단원8점수 4개	기말시험점수
C3	단원1점수~단원4점수 4개	중간시험점수
C4	단원1점수~단원8점수 8개	최종석차

와 KNN 모델의 테스트 결과만 분석하였다. 전체 31명 학생 그룹 분석에서는 Data Sampler 위젯을 통해 31개 학습데이터 중 실제 모델 학습에 사용된 데이터 수는 22개, 테스트용 데이터 수는 9개로 설정하였다. 표 6부터 표 8은 Predictions 위젯에서 2개 모델을 테스트한 RMSE 결과와 2개 모델의 RMSE 평균값, 모델의 예측정확도 평균값을 나타낸다. 이 4개 값들이 C1~C4 4개 시나리오 별로 도출되었다. C1~C4 4개 시나리오에서 사용된 목적변수가 모두 100점 만점이므로, 100에서 RMSE 평균값을 뺀 값을 모델의 예측정확도 평균값으로 계산하였다.

상위성적 15명 학생 그룹 분석에서는 Data Sampler 위젯을 통해 15개 학습데이터 중 실제 모델 학습에 사용된 데이터 수는 11개, 테스트용 데이터 수는 4개로 설정하였고, 하위성적 16명 학생 그룹 분석에서는 Data Sampler 위젯을 통해 16개 학습데이터 중 실제 모델 학습에 사용된 데이터 수는 12개, 테스트용 데이터 수는 4개로 설정하였다. 표 6부터 표

표 6. 전체그룹 최종성적 예측 머신러닝 모델 테스트 결과

Table 6. Prediction test results of final grade machine learning model for total group

시나리오	RMSE-C1	RMSE-C2	RMSE-C3	RMSE-C4
Random Forest	21.3	23.7	20.3	9.0
kNN	19.6	23.9	20.6	7.6
RMSE평균	20.5	23.8	20.4	8.3
예측정확도(%)	79.5	76.2	79.6	91.7

표 7. 상위그룹 최종성적 예측 머신러닝 모델 테스트 결과

Table 7. Prediction test results of final grade machine learning model for top group

시나리오	RMSE-C1	RMSE-C2	RMSE-C3	RMSE-C4
Random Forest	8.6	13.8	5.1	4.3
kNN	5.6	14.1	5.1	3.2
RMSE평균	7.1	14.0	5.1	3.8
예측정확도(%)	92.9	86.0	94.9	96.2

표 8. 하위그룹 최종성적 예측 머신러닝 모델 테스트 결과

Table 8. Prediction test results of final grade machine learning model for bottom group

시나리오	RMSE-C1	RMSE-C2	RMSE-C3	RMSE-C4
Random Forest	4.3	17.6	9.3	2.1
kNN	9.1	17.8	5.9	1.7
RMSE평균	6.7	17.7	7.6	1.9
예측정확도(%)	93.3	82.3	92.4	98.1

8의 결과로부터 전체 31명에 대한 예측 정확도보다 상위성적 그룹 15명과 하위성적그룹 16명에서 높은 예측 정확도를 나타내었다. 이러한 경향은 모든 예측 시나리오 C1~C4에서 모두 나타났다.

IV. 예측 시나리오별 모델 예측 정확도 분석

표 9에서 표 13은 표 5에서 설명한 C1~C4 예측 시나리오별로 정리하여, 머신러닝 모델의 예측 정확도를 나타내었다. 전체 그룹 보다 상위 또는 하위 성적 그룹으로 나누어 단원별 연습문제 점수 데이터 8개로부터 최종성적을 예측하는 것이 정확도가 높았다. 그리고 C2 시나리오, 즉, 단원5점수~단원8점수 4개로부터 기말시험점수를 예측하는 시나리오의 정확도가 가장 낮았고, C4 시나리오, 즉, 단원1점수~단원8점수

표 9. C1 시나리오의 최종성적 예측에 대한 그룹별 결과

Table 9. Results by group for prediction of final grade in C1 scenario

그룹	전체31명	상위15명	하위16명	평균값
RMSE평균	20.5	7.1	6.7	11.4
예측정확도(%)	79.5	92.9	93.3	88.6

표 10. C2 시나리오의 최종성적 예측에 대한 그룹별 결과

Table 10. Results by group for prediction of final grade in C2 scenario

그룹	전체31명	상위15명	하위16명	평균값
RMSE평균	23.8	14.0	17.7	18.5
예측정확도(%)	76.2	86.0	82.3	81.5

표 11. C3 시나리오의 최종성적 예측에 대한 그룹별 결과

Table 11. Results by group for prediction of final grade in C3 scenario

그룹	전체31명	상위15명	하위16명	평균값
RMSE평균	20.4	5.1	7.6	11.0
예측정확도(%)	79.6	94.9	92.4	89.0

표 12. C4 시나리오의 최종성적 예측에 대한 그룹별 결과

Table 12. Results by group for prediction of final grade in C4 scenario

그룹	전체31명	상위15명	하위16명	평균값
RMSE평균	8.3	3.8	1.9	4.7
예측정확도(%)	91.7	96.2	98.1	95.3

표 13. 시나리오별 모델의 최종 성적 예측 정확도

Table 13. Final grade prediction accuracy of model by scenario

시나리오	C1	C2	C3	C4
예측정확도(%)	88.6	81.5	89.0	95.3

8개로부터 최종석차를 예측하는 시나리오의 정확도가 가장 높았다.

표 12의 C4 시나리오는 그룹별 결과에서 모두 90%이상 높은 예측 정확도를 나타내었다. 이에 따라, 파이썬 단원별 연습문제 점수들로부터 학생의 최종 성적을 예측하는 머신러닝 모델의 시나리오는 C4로 선정할 수 있다. 이러한 C4 시나리오로 분석했을 때, 단원별 자동채점 연습문제의 출제 타당도가 높았음을 알 수 있다.

V. 결 론

본 논문에서 제안한 파이썬 프로그래밍 단원별 연습문제의 타당성 분석 방법은 파이썬 자동 채점 연습 문제와 머신러닝 모델 구축 워크플로우, 그룹별 모델 예측 정확도 분석, 예측 시나리오별 모델 예측 정확도 분석으로 구성되었다. 제안한 단원별 연습문제의 타당성 분석 방법을 통해 자동채점 연습문제들의 타당도를 분석하여 단원별 연습문제들을 개선할 수 있다. 본 논문에서는 단원별 연습문제 점수로부터 최종성적 예측 모델의 정확도를 산출하였고, 이러한 정확도는 단원별 연습문제의 타당성과 비례한다고 가정하였다. 특히, 최종 성적의 석차를 예측하는 시나리오의 정확도가 가장 높게 나와 이러한 시나리오를 추천하였다. 모든 실습 과목에서 이러한 연습문제들의 개선은 필요하며, 본 논문에서는 최종 성적과 연계하여 종합적으로 분석하는 프레임을 제안하였다. 더 많은 데이터를 확보하여 모델의 정확도를 산출할 필요가 있으며, 파이썬 단원별 연습문제 점수들로부터 학생의 최종 성적을 예측하는 머신러닝 모델의 예측 정확도로부터 단원별 자동채점 연습문제의 출제 타당도를 분석할 수 있음을 결과로 제시하였다.

참고문헌

[1] J. K. Jeong, “Design and construct of programming assessment system based on online judge for a science high school student,” Korea National University of Education, 2010.

[2] Codle, Team Monolith, 2022, [Online]. Available: https://codle.io/.

[3] W. Y. Chang and S. S. Kim, “Development and application of algorithm judging system : analysis of effects on programming learning,” *The Journal of Korean Association of Computer Education*, vol. 17, no. 4, pp. 45-57, 2014.

[4] Orange, University of Ljubljana, 2023, [Online]. Available: https://orangedatamining.com/.

[5] S. S. Kim, S. H. Oh, and S. S. Jeong, “Development and application of problem bank of problem solving programming using online judge system in data structure education,” *The Journal of Korean Association of Computer Education*, vol. 21, no. 4, pp. 11-20, 2018.

[6] S. S. Jeong, “The effects of programming education using an automatic programming assessment system on learning flow of general high school students,” Korea National University of Education, 2019.

[7] W. Y. Chang, “The effects of online judge system in programming education on learning motivation and thinking : structural relationships between factors,” Korea National University of Education, 2020.

[8] K. Hur, “Python basic programming curriculum for non-majors and development analysis of evaluation problems,” *Journal of Practical Engineering Education*, vol. 14, no. 1, pp. 75-83, 2022.

허 경 (Kyeong Hur)_종신회원

1998년 고려대 전자공학과 학사
2000년 고려대 전자공학과 석사
2004년 8월 고려대 전자공학과 통신공학박사
2004년 8월~2005년8월 삼성종합기술원(SAIT) 전문연구원
2005년 9월~현재 경인교대 컴퓨터교육과 교수
〈관심분야〉 네트워크 MAC QoS, IoT, SW교육, AI교육, 데이터과학교육