

전공자와 비전공자 간의 교양과목 평가 분리 여부 결정을 위한 분석 시스템에 관한 연구[†]

오다솜*, 최보아*, 김주은**, 이종혁*‡

*대구가톨릭대학교 빅데이터공학과

**대구가톨릭대학교 제약공학과

e-mail : jonghyuk@cu.ac.kr

A Study on the Analysis System for Determination of Separation of Liberal Arts Evaluation between Majors and Nonmajors

DaSom Oh*, BoAh Choi*, Joo-Eun Kim**, JongHyuk Lee*

*Dept. of Big Data Engineering, Daegu Catholic University

**Dept. of Pharmaceutical Engineering, Daegu Catholic University

요 약

대학기관연구는 대학에서 이루어지는 다양한 의사결정을 체계적으로 지원하는 활동이다. 본 대학에서는 최근에 전공자와 비전공자 간의 교양과목 평가 분리에 대한 제도를 마련하였으나 이 결정을 교수자의 재량에 맡겨 놓아 근거 데이터 없이 교수자가 평가 분리 여부를 실제로 결정하기에는 어려운 실정이다. 이에 본 논문은 이전 학기의 성적 데이터를 기반으로 통계 분석한 결과와 이에 대한 시각화를 제공하여 교수자가 조금 더 쉽게 평가 분리 여부를 결정할 수 있도록 분석 시스템을 제안하고 실제 교양과목에 대한 분석 결과를 예시를 통해 보여준다.

1. 서론

대학기관연구(Institutional Research)는 대학에서 이루어지는 다양한 기획 및 정책개발 그리고 이와 관련된 의사결정을 체계적으로 지원하기 위한 일체의 활동을 의미한다[1]. 대학기관연구 분야 중 대학이 당면한 문제를 데이터 분석을 통해 해결하는 분야도 해당된다. 최근 본 대학에서는 전공과목 중 교양 수준의 내용으로 과목 개설을 요구하는 수요가 있어 많은 교양과목이 개설되고 있다. 하지만 성적 평가 후에 교양과목의 난이도 등 여러 이유로 학생들로부터 전공자와 비전공자 간 성적에 대한 형평성 문제를 제기하여 결국 성적 평가를 분리해달라는 요구가 있어 왔다.

이에 본 대학은 강의 계획 시 교수의 재량에 따라 전공자와 비전공자 간의 성적 평가를 분리할지 여부를 결정할 수 있도록 하였다. 그러나 현재 시스템은 교양과목 교수자에게 평가 분리 여부 결정을 전적으로 맡겨 놓았기 때문에 교수자가 데이터 분석 결과를 기반으로 결정에 도움을 줄 수 있도록 하는 시스템 개선이 필요하다. 본 논문은 전공자와 비전공자 간의 성적 분리 평가 여부를 데이터 분석 결과를 기반으로 결정하기 위한 시스템 아키텍처를 제안하고 실제로 이전에 개설된 과목의 성적 데이터를 바탕으로 독립 표본 검정(t-test)을 이용해 전공자와 비전공자 간의 성적 차이가 발생하는지에 대해 검증한다.

본 논문의 2 절에서는 전공자와 비전공자 간의 성적 차이 존재 여부를 확인하기 위해 이용한 독립표본 t 검정을 설명하고 3 절에서는 본 논문이 제안하는 분석 시스템 아키텍처를 설명한다. 다음 4 절에서는 실제 데이터를 이용하여 전공자와 비전공자 간의 성적 차이가 있는지 통계 분석한 결과를 설명한다. 마지막으로 5 절에서는 결론을 맺는다.

2. 관련연구

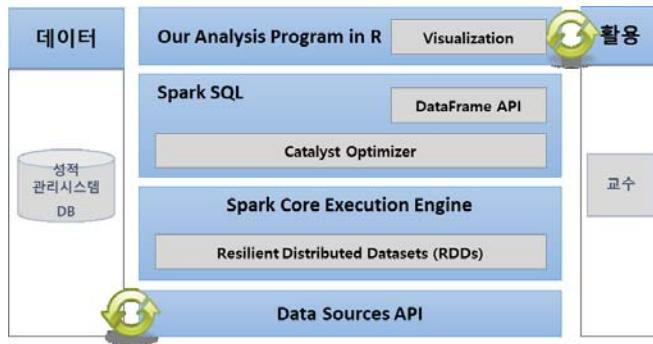
t 검정[2]은 t 분포에 의존하여 두 집단 간의 평균이 통계적으로 유의미한 차이를 보이는지 판별한다. 본 논문은 전공자와 비전공자의 성적 차이를 분석하기 위해 독립표본 t 검정(independent samples t-test)을 사용한다. 독립표본 t 검정은 두 개의 서로 다른 모집단의 평균에 대한 차이가 있는지를 검정하기 위하여 각각의 모집단에서 추출된 서로 독립적인 두 개의 표본의 평균값을 비교하고, 이를 통해 모집단 간의 차이가 있는지를 확인하는 통계적 기법이다. 본 논문에서 사용한 종속변수인 성적은 비율 척도이며 연속형 자료이고 독립변수인 전공 유무와 학년은 명목 척도이며 범주형 자료이다. 따라서 연속형 자료와 범주형 자료 간의 통계분석에 적합한 기법인 독립 표본 t 검정을 본 논문의 분석 방법으로 사용한다.

[†] 본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW 중심대학지원사업의 연구결과로 수행되었음(2019-0-01056).

[‡] 교신저자

3. 분석 시스템 아키텍처

본 논문은 빅데이터 애플리케이션 개발에 필요한 통합 플랫폼으로서 스파크(Spark) 프레임워크를 이용한다. (그림 1)은 성적 분리 평가 여부에 대한 결정을 도와주기 위한 분석 시스템 아키텍처이다. 그림에서 보듯이 학내 성적 관리 시스템에 있는 데이터소스에 접근하기 위해 스파크의 데이터소스 API를 사용하며 이를 스파크 SQL로 쿼리하고 R 언어를 이용해 성적 데이터를 통계 분석한다. 그리고 통계 분석한 결과를 시각화하여 교수자가 활용할 수 있도록 한다.



(그림 1) 분석 시스템 아키텍처

4. 분석 과정 및 결과

본 논문은 전공자와 비전공자 간 성적 차이가 있는지 데이터를 분석하기 위해 본 대학에서 2018년 2학기에 개설한 교양과목인 클라우드컴퓨팅의 이해와 기초빅데이터분석 과목의 성적 데이터를 수집하였으며 익명화, 결측치 처리 등과 같은 데이터 정제를 하였다. 클라우드컴퓨팅의 이해 과목은 클라우드 컴퓨팅의 기본 이론을 학습하고 Amazon Web Services를 실습하여 클라우드 컴퓨팅을 이해하도록 하는 내용으로 구성되어 있다. 그리고 기초빅데이터분석 과목은 R 언어의 기초 문법을 학습하고 간단한 데이터를 분석해 보는 내용으로 구성되어 있다. 과목 내용에서 알 수 있듯이 이 두 과목은 컴퓨터 전공 관련 과목이다. 본 논문은 이 두 과목을 수강한 학생의 학부 및 학과 이름을 통해 다음과 같이 전공자와 비전공자로 분류하였다.

- 전공자: IT 공학부, 빅데이터공학과
- 비전공자: 그 외 나머지 학과

본 논문은 전공자와 비전공자의 성적에 대한 차이를 알아보기 위해 다음과 같이 가설을 세우고 독립표본 t 검정을 실행하여 분석하였다.

- H_0 : 전공자와 비전공자 간의 성적에 대한 평균 차이가 없다.
- H_1 : 전공자와 비전공자 간의 성적에 대한 평균 차이가 있다.

4.1 클라우드컴퓨팅의 이해 과목 성적 차이 분석

클라우드컴퓨팅의 이해 과목은 <표 1>과 같이 총 41명이 수강하였다. 독립표본 t 검정 수행 결과, <표 2>와 같이 비전공자 24명의 성적에 대해 비전공자 평균은 65.21739이고, 전공자 18명의 성적 평균은 76.05556이다. 유의수준 $\alpha=0.05$ 에서 집단간의 평균에 대한 유의확률을 검정하였을 때, 유의확률은 0.03264이므로 H_1 을 채택한다. 따라서 전공자와 비전공자간의 성적에 대한 평균 차이가 있다고 할 수 있다.

<표 1> 클라우드컴퓨팅의 이해 과목 수강 인원

	전공자	비전공자
1 학년	12	16
2 학년	2	2
3 학년	3	3
4 학년	1	2
합계	18	23

<표 2> 클라우드컴퓨팅의 이해 과목 분석 결과

	성적 평균	유의확률
전공자	76.05556	0.03264
비전공자	65.21739	

4.2 기초빅데이터분석 과목 성적 차이 분석

기초빅데이터분석 과목은 <표 3>과 같이 총 42명이 수강하였다. 독립표본 t 검정 수행 결과, <표 4>와 같이 비전공자 26명의 성적에 대해 비전공자 평균은 67.34783이고, 전공자 16명의 성적 평균은 63.60000이다. 유의수준 $\alpha=0.05$ 에서 집단간의 평균에 대한 유의확률을 검정하였을 때, 유의확률은 0.6811이므로 H_0 을 채택한다. 따라서 전공자와 비전공자 간의 성적에 대한 평균 차이가 없다고 할 수 있다.

<표 3> 기초빅데이터분석 수강 학생

	전공자	비전공자
1 학년	6	8
2 학년	2	9
3 학년	3	5
4 학년	5	4
합계	16	26

<표 4> 기초빅데이터분석 과목 분석 결과

	성적 평균	유의확률
전공자	63.60000	0.6811
비전공자	67.34783	

위의 두 과목 분석 결과에 따르면 클라우드컴퓨팅의 이해 과목은 전공자와 비전공자 간의 성적에 대한 평균 차이가 있다. 하지만, 기초빅데이터분석 과목은 전공자와 비전공자 간의 성적에 대한 평균차이가 없다. 클라우드컴퓨팅의 이해 과목 데이터 분석 결과에 대한 요인이 무엇인지 파악하고자 특징에 대해 검토해본 결과, 수강 전체 인원 중 1 학년 학생의 비율이 모든 강의에서 가장 높은 것을 알 수 있었다. 즉, 1 학년의 성적이 분석에서 가장 높은 영향력을 가지고 있다. 교육 과정상 1 학년 때는 전공 과목을 이수하지 않거나 적은 수의 전공 강의를 수강한다. 이는 1 학년 학생을 전공자와 비전공자로 분류하는 것이 적절한지에 대한 문제로 이어져 다음과 같은 가설을 세우고 두 과목의 1 학년 학생들에 대한 전공자와 비전공자 간의 성적 차이가 있는지 독립표본 t 검정을 실행하여 분석하였다.

- H_0 : 1 학년 전공자와 비전공자 간의 성적에 대한 평균 차이가 없다.
- H_1 : 1 학년 전공자와 비전공자 간의 성적에 대한 평균 차이가 있다.

4.3 클라우드컴퓨팅의 이해 과목 1 학년 성적 차이 분석

클라우드컴퓨팅의 이해 과목은 <표 1>과 같이 1 학년 28 명이 수강하였다. 독립표본 t 검정 수행 결과, <표 5>와 같이 비전공자 12 명의 성적에 대한 비전공자 평균은 67.5625이고, 전공자 평균은 80.0000이다. 유의수준 $\alpha=0.05$ 에서 집단간의 평균에 대한 유의확률을 검정 하였을 때, 유의확률은 0.008465 이므로 H_1 을 채택한다. 따라서 1 학년 전공자와 비전공자 간의 성적에 대한 평균 차이가 있다고 할 수 있다.

<표 5> 클라우드컴퓨팅의 이해 과목 1 학년 분석 결과

	성적 평균	유의확률
전공자	80.0000	0.008465
비전공자	67.5625	

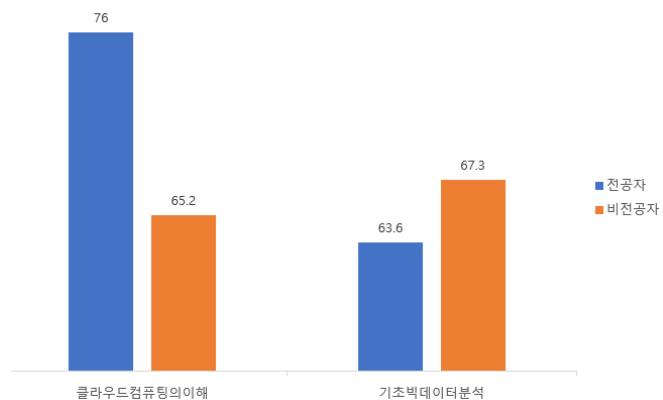
4.4 기초빅데이터분석 과목 1 학년 성적 차이 분석

기초빅데이터분석 과목은 <표 3>과 같이 1 학년 14 명이 수강하였다. 독립표본 t 검정 수행 결과, <표 6>과 같이 비전공자 14 명의 성적에 대한 비전공자 평균은 49.66667이고, 전공자 평균은 56.16667이다. 유의수준 $\alpha=0.05$ 에서 집단간의 평균에 대한 유의확률을 검정 하였을 때, 유의확률은 0.3466 이므로 H_0 을 채택한다. 따라서 1 학년 전공자와 비전공자 간의 성적에 대한 평균 차이가 없다고 할 수 있다.

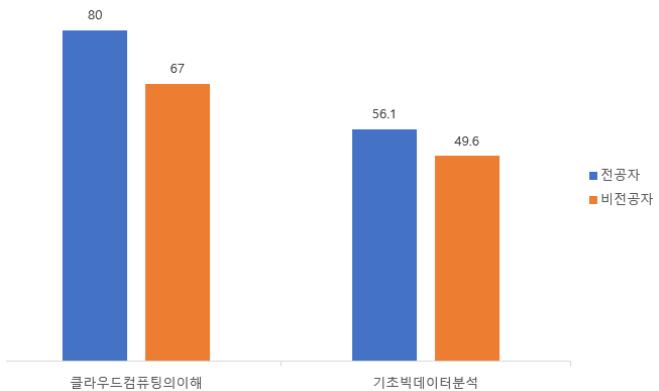
<표 6> 기초빅데이터분석 과목 1 학년 분석 결과

	성적 평균	유의확률
전공자	56.16667	0.7482
비전공자	49.66667	

위의 두 과목의 1 학년 학생들에 대한 분석 결과에 따르면 전학년을 대상으로 한 분석 결과와 다르지 않다. 즉 클라우드컴퓨팅의 이해 과목은 전공자와 비전공자 간의 성적에 대한 평균 차이가 있는 반면 기초빅데이터분석 과목은 차이가 없다. (그림 2)는 전학년 대상으로 전공자와 비전공자의 성적 평균을 보여준다. 그림에서 보듯이 클라우드컴퓨팅의 이해 과목에서 평균 차이가 크지만 기초빅데이터분석 과목은 작음을 알 수 있다. (그림 3)은 1 학년 대상으로 전공자와 비전공자의 성적 평균을 보여준다. 그림에서 보듯이 클라우드컴퓨팅의 이해 과목이 기초빅데이터분석 과목보다 성적 평균 차이가 큼을 알 수 있다. 본 논문은 교수자에게 이와 같은 통계분석 결과(성적 평균, 유의확률 등)와 시각화 결과(성적 평균 비교 그래프 등)를 제공할 계획이다. 본 논문에서 사용한 성적 데이터 및 분석 프로그램은 GitHub[4]에 공개하였다.



(그림 2) 전학년 대상 전공자와 비전공자의 성적 평균



(그림 3) 1 학년 대상 전공자와 비전공자의 성적 평균

5. 결론

본 논문은 교수자가 교양과목 평가 분리 여부를 데이터 기반으로 결정할 수 있도록 분석 시스템을 제안하였고 실제 성적 데이터 분석을 통해 교양과목별 전공자와 비전공자 간의 성적 차이가 존재하는지에 대해 알아보았다. 분석 결과, 특정 과목에서는 성적 차이가 있었으나 성적 차이가 없는 과목 또한 있었다. 교수자에게 강의 계획 시 과거 성적 데이터 분석 결과와 그래프 등의 시각화를 제공하여 전공자와 비전공자 간의 평가 분리 결정에 도움을 준다면 향후 학생의 성적 평가 형평성 문제 제기가 줄어들 것으로 예상된다.

본 논문은 계속해서 본 논문에서 제안한 방법이 본 대학의 학사 시스템에 적용되도록 컴퓨터 전공 관련 교양과목 외에 다양한 교양과목 평가 분석을 통해 현재 개발 결과를 보완할 계획이다.

참고문헌

- [1] 배상훈, 윤수경, “한국대학에서 대학기관연구 (Institutional Research) 도입 및 관련 쟁점과 시사점,” 아시아교육연구, 17 권 2 호, 2016.
- [2] Student, “The Probable Error of a Mean,” Biometrika, 6 권 1 호, 1908.
- [3] Apache Spark, <https://spark.apache.org/>
- [4] https://github.com/boahchoi/KIPS_2019.git