

빅데이터를 접목한 스마트시대 온라인 학습 모델의 제안과 실증

박재천¹ · 이두영^{2*} · 국성희³

Proposal of Smart era Online Learning Model with BigData

Jae-chun Park¹ · Doo-young Lee^{2*} · Sung-hee Kuk³

¹Department of IT&Media Convergence Studies, INHA University, Incheon 402-751, Korea

^{2*}Department of IT&Media Convergence Studies, INHA University, Incheon 402-751, Korea

³Department of Statistics, INHA University, Incheon 402-751, Korea

요 약

본 논문은 스마트시대의 온라인 학습에 대한 논문으로, 새로운 모델을 제안하고 실증하는데 초점을 두었다. 온라인 학습 클래스 운영에 있어 각 학습 요인들을 통해서 최종 성취도를 예측하는 연구를 진행하였다. 이에 학습 운영 요인 7가지를 정하고 학습자들의 데이터를 수집한 후 의사결정나무방법을 통한 예측 모델을 완성한다. 모델을 통한 예측성을 확인한 후, 일반성 확보를 위해 다른 교과목에도 모델을 적용시켜 예측성을 확인하였다. 결과적으로 기존의 온라인 클래스의 정적인 학습 모델을 넘어 객관적인 지표를 이용한 학업성취도를 상시적으로 확인할 수 있게 하였다. 학습자와 교수자 모두가 학습 중 유용하게 활용할 수 있는 스마트시대 새로운 패러다임의 학습 모델을 제안한다.

ABSTRACT

This paper is studying for New Online Learning Model on Smart digital era. It can expect the result of learning degree on Online Learning Class. Using 7-factors of Online Class's operating policy, make the expectation model by 'decision tree' method. And through applying another class, we can getting a generality. Finally, Over the traditional Online Class model, we can take the real-time status of Online class learning degree. It is useful both students and teacher. It is the one of 'Smart learning Model'.

키워드 : 스마트러닝, 이러닝, 빅데이터

Key word : Smart Learning, e-learning, BigData

Received 03 December 2014, Revised 21 December 2014, Accepted 05 January 2015

* Corresponding Author Doo-Young Lee(E-mail:leedy87@inha.edu, Tel:+82-32-860-8498)

Department of IT&Media Convergence Studies, INHA University, Incheon 402-751, Korea

Open Access <http://dx.doi.org/10.6109/jkiice.2015.19.4.991>

print ISSN: 2234-4772 online ISSN: 2288-4165

©This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서 론

최근 정보통신 인프라와 모바일 디바이스의 급격한 발달로 인하여 전사회적으로 생활패턴의 변화가 일고 있다. 공간을 초월한 정보공유에서 나아가 모든 사물들을 연결하려는 이른바 ‘초연결’ 사회에 우리는 진입하고 있다. 덕분에 우리는 전에 없던 정보 호황을 누리고 있고, 각종 산업군이나 여러 방면에서 업무나 생활의 효율성이 증대되고 있다. ‘스마트 시대’라고 불리는 시대이다. 스마트 시대의 독특한 특징으로 인하여 각종 이슈들이 대두되고 있고, 이전에는 불가능하던 일들이 가능하게 되었다[1].

본 연구에서는 스마트 시대의 주요 특징이 이러닝 산업에 어떻게 작용할 수 있는지에 대한 모형을 제시하고 통계적 분석을 통해 실증하고 있다. 다루는 키워드로는 스마트 시대의 특징 중 ‘빅데이터’ 모델에 집중하였고, 이런 현상이 ‘온라인 학습’에 어떻게 접목이 가능할 수 있는지에 대한 연구를 진행하였다. 각종 기관이나 학교, 회사 등에서 ‘이러닝’이라고 불리는 온라인 학습이 활발히 사용되고 있다. 이러닝은 언제 어디서나 무엇이든 학습할 수 있는 환경을 포괄적으로 지칭하고 있으며, 그 효율성이 높아 그 활용도가 높아지고 있다[2]. 이에 국가 차원에서 이 이러닝 산업을 정책적으로 지원하고 있다. 하지만 이러닝이 ‘교육’의 카테고리 내에서 다루어져야 함에도 불구하고, 경제적, 비용적 차원에서 이용되는 현실적 한계도 존재한다[3].

따라서 본 논문에서는 빅데이터가 온라인 학습에 어떻게 적절하게 적용될 수 있는지 모형을 설계하고 실제 대학의 이러닝 클래스 현장에 적용하여 분석한다. 학습자들의 수강기록 빅데이터를 기반으로 하여 학습자의 현황에 대해서 객관적으로 파악할 수 있는 모델을 제안한다. 또한 본 모델이 일반적으로 적용될 수 있는지 확인하기 위하여 다른 클래스에도 적용하는 방법을 통하여 일반성을 확보하고 있다. 이런 타입의 온라인 학습 모델을 ‘스마트 러닝’이라고 부르기로 한다.

II. 본 론

2.1. 스마트 시대의 특성

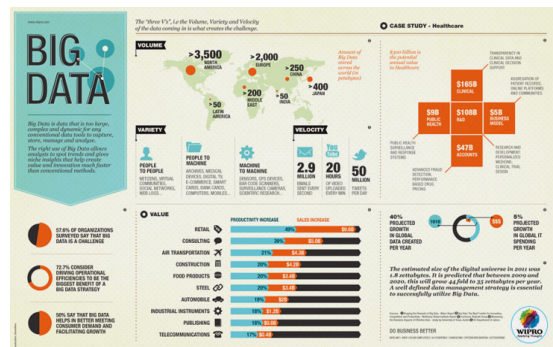
스마트 러닝을 논의하기에 앞서, 스마트 시대란 구

체적으로 어떤 현상을 보이는 배경인지를 파악할 필요가 있다.

우선, ‘인터넷+폰’이라는 개념으로 탄생한 아이폰을 효시로 하여 스마트폰을 필두로 한 스마트 디바이스의 활황기라고 볼 수 있다. 여기서 스마트 디바이스는 TCP/IP 프로토콜을 사용하는 인터넷에 연결된 모바일 디바이스로, 태블릿, 스마트 TV, 스마트폰 등을 예로 들 수 있다. 또한 네트워크 인프라와 무선기술의 발달이 뒷받침 되어 이전 세대에서 접하기 어려웠던 고용량 멀티미디어 콘텐츠를 모바일 디바이스를 통해 언제 어디서나 수월하게 접할 수 있게 된다.

둘째로, 무한정에 가까운 양의 데이터가 생성되고 축적되고 있는 현상이다. 정보 생산자가 스마트 디바이스를 통해 온라인 공간에서 엄청난 양의 정보들이 생성하고 그 정보들을 재생산하고 있다. 여기서 데이터는 사람이 생성해 내는 데이터를 넘어 사물들에 장착된 센서에 의해서 생성된 데이터도 포함된다. 사물 간에도 네트워크가 만들어져 통신을 하는데, 최근 이 개념이 ‘Internet of Things’, 사물인터넷이라고 불리우는 개념이다[4].

이렇듯 사람과 사물 할 것 없이 정보들을 생산해내고 재구성해 내는 과정에서 ‘유의미한’ 정보가 탄생한다. 이를 유용하여 수많은 데이터들을 토대로 하여 유의미한 정보, 즉 ‘지식’을 만들어 낼 수 있는 환경은 ‘빅데이터’라고 불리우며, 스마트 시대에서 나타나는 현상 중 하나이다.



Source: <http://cfite21.uf.tistory.com/image/17500F3351429CBC3804F5>

그림 1. 빅데이터 활용도 관련 타이포그래피
Fig. 1 BigData web typography

다음으로, Web2.0 개념으로 통용되는 온라인 공간에서의 양방향적 정보흐름과 소셜 네트워킹 시스템의 발달로 촉발된 온라인 공간에서 커뮤니케이션 패턴의 변화다. Web2.0 개념은 'O'Reilly Media'에서 2004년부터 사용하기 시작한 개념으로, 미국 IT버블에서 살아남은 인터넷 기업들의 공통점을 아우르는 개념이다. 이는 단방향적인 정보흐름을 탈피하고 양방향 적인 정보흐름을 지향한다[5].

이는 누구나 정보 생성자인 동시에 정보 수용자가 되는 것이고, 이는 서비스 이용자들의 자발적인 참여를 통해 구현되게 된다. 이후 web2.0 개념은 온라인 공간에서의 기본적인 패러다임으로 자리잡게 되었다. 이와 더불어 커뮤니케이션 패턴 또한 변화하게 된다. 소셜 네트워킹 시스템의 활성화로 단방향적 정보전달 방식에 효율적이었던 기존의 '게시판' 형식에서 'Micro Bloging' 내지 'Mention' 등의 방식으로 정보 전달 방식이 민첩하게 바뀌었고, 'comment' 혹은 '추천' 방식의 피드백을 통해 즉각적이고 신속한 커뮤니케이션 패턴으로 변모하게 되었다[6].



Source: twitter.com

그림 2. SNS의 커뮤니케이션 패턴
Fig. 2 Communication Pattern of SNS

스마트 시대란 '인터넷에 연결된 스마트 디바이스를 통해 사람과 사물이 연결된 세상이며, 생성되는 데이터의 양은 무한대에 가깝고 모든 Peer들이 활발히 네트워킹에 참여하고 있는 세상'이라고 정리할 수 있다. 따라서 스마트 시대의 온라인 학습 패턴 역시 시대적인 추세를 반영하는 적합한 모델이 필요한 것이다.

2.2. 온라인 학습 환경의 발전[4]

온라인 학습의 효시는 전통적으로 '집합교육'과 '원격교육'의 대비에서 나타났다고 볼 수 있다. 인간사회에 있어서 '학습'이란 어느 시대나 존재해 왔고, 학습에 있어 시간과 공간의 한계를 극복하기 위한 노력이 많았다. 우편 교육이 그랬고, 이는 방송교육으로 변화하고 곧 이러닝으로 변화하게 된다. 그 기저에는 정보통신기술의 발전이 깔려있다. 따라서 온라인 학습의 발전사는 정보통신기술의 발전사와 맞물려 있는 점들이 많다.

컴퓨터를 본격적으로 학습에 이용하게 된 시절의 온라인 학습을 Electronic Learning, 전자학습 혹은 이러닝 시대라고 본다. 이러닝으로 IT기술과 컴퓨터를 학습에 활용하여 언제, 어디서나, 누구나, 수준별 학습이 가능한 학습모델이 탄생하였다. 그 이후에 등장한 온라인 학습 모델은 '혼합형 학습'이라고 불리는 블렌디드러닝(Blended Learning), B러닝이다. B러닝은 이러닝의 한계점을 보완하고 장점을 취합하기 위해 온라인, 오프라인 혼합 학습 모델을 제시하였다.

이후 모바일 디바이스의 발전으로 모바일러닝(Mobile Learning), M러닝 개념이 등장하였다. 무선 디바이스를 통해 학습 환경 자체가 Universe 하게 확장된다. 모바일 디바이스는 책상 앞에서 벗어나 학습자가 위치한 곳 어디서나 학습을 가능하게 하였다. 이후 Ubiquitous Learning, 유러닝은 센서와 네트워킹 기술의 발달로 일상생활에서 접하는 모든 상황에서 학습이 가능한 환경을 제시해 주었다.

스마트 시대의 온라인 학습, 스마트 러닝 모델은 이전 모델 보다 더 나은 기술을 접목시킨다기보다는, 주어진 환경에서 조금 더 효율적이고 창의적인 발상의 전환이 필요하다. 이른바 '패러다임'의 전환을 통한 학습 모델을 요하고 있다.

2.3. 기존 온라인 학습 모델의 한계와 시사점

기존 대학에서는 온라인 학습, 즉 '이러닝'이라고 불리는 클래스가 활발히 이용되고 있다. 인천의 A대학에는 수십개의 이러닝 강좌가 개설되어 있다. 전통적인 방식의 오프라인 강의와 온라인 강의의 가장 큰 차이점은, 수강생 수에 있다. 일반적인 경우, 오프라인 강의의 수강생은 수십명 내외인 것에 비해, 이러닝 강의의 경우 수백명의 학생이 수강하는 경우가 빈번하다. 이는

이러닝의 장점으로 꼽히는 시간과 공간의 제약에서 벗어나는 것 이외에 다른 이유도 존재한다.

우선 이러닝 운영 주체인 대학의 입장에서는, 단기에 만들어진 강의 콘텐츠를 통해 수년에 걸쳐 활용할 수 있기에 강의자 확보에 있어서 경제적이라 볼 수 있다. 또한 물리적인 강의실이 필요없어 공간부족 문제에서도 해방되어 이러닝은 운영측면에 있어서 경제성이 높다는 장점이 있다.

학생의 입장에서 이러닝은 수강에 효율적이라는 장점을 얻는다. 물리적인 강의실에 출석할 필요가 없다는 점은 상당한 시간과 비용을 절약하는 효과가 있다. 또한 비교적 정적인 이러닝 학습 내용으로 인하여 오프라인 강의에 비하여 학기 중 비교적 부담없이 수강하고 원하는 학점을 취득할 수 있기에 효율적이라고 할 수 있다.

이러닝 운영의 이런 실제적인 현상에는 몇가지 부작용이 존재한다. 학습 공급자인 교수자의 물리적 필요 역량이 줄어들고, 학습 수요자인 학생들의 수강 수요가 늘어나는 것은 전체적 구조로 보았을 때 경제적이라 할 수 있겠다. 하지만 비대면적인 특성을 가진 이러닝 클래스에서 한정된 조건을 가진 교수자가 수많은 수강생들을 일일이 체크하고 학습에 몰입시키기엔 한계가 존재한다. 따라서 클래스 운영 모델이 단순화 될 수 밖에 없고, 결국 고착화된 시스템이 구축이 되버린다. 학습이라는 재화는 그 중요성이 경제적 논리에 의해 다루어져서는 안되는데도 불구하고 교육의 운영양상이 변화하고 있는 것이 현실이다.

우리 사회에 온라인 학습이 널리 활용되고 있고 발전이 되는 상황이다. 교육의 본질에 대해서 심대하게 다루는 것은 나중으로 미루더라도, 우선은 주어진 환경들을 우리의 학습을 위해서 어떻게 유용하고 효과적으로 사용할 수 있는 것인가를 주지해야 한다. 시대적 흐름은 ‘스마트 시대’에서 더욱 미래로 치닫고 있는데, 실제 교육 현장에서 이루어지는 온라인 학습은 초기 모델인 ‘이러닝’ 수준에 머물러 있는 것이 현실이다. 따라서 새로운 온라인 학습 모델에 대한 요구는 언제나 있어 왔고, 연구들이 있었으나 현장에서 바로 적용할 수 있는 구체적 모델은 부족하다. 또한 차세대 온라인 학습 모델에 대한 이론적 연구에 머물러 있어 실증적인 연구에 대한 필요가 있는 것이다.

Ⅲ. 대안 모델 제시 ; Data based Getting Position Method, DPM

온라인 학습 모델의 ‘스마트’한 운영방식 전환을 위한 첫걸음으로 ‘지피지기’ 자세가 필요하다. 온라인 학습을 한학기동안 수강하는 학생의 입장에서는, 한학기 내내 다른 수강생들과 접촉할 기회가 적다. 이러닝 클래스 웹 사이트에서 게시판 형식을 통해 질의 응답 정도가 수강생들 간에 접촉할 수 있는 정도이다. 따라서 현재 수강생 본인이 수업에 잘 따라가고 있는지, 상대적인 위치가 어느정도 되는지, 어떠한 부분을 잘하고 있고 어떠한 부분이 부족한지를 파악하기 어렵다. 교수자 입장에서도 수백명의 온라인 클래스 수강생들의 출결 정보나 시험점수 등 숫자화된 정보로 면밀한 수강생 파악하는 데에 한계가 있다. 따라서 수강기간 중 학습자들의 상대적인 위치나 장단점을 파악할 수 있는 지표가 있다면, 온라인 클래스에 있어서 수강생과 교수자 입장 모두에 도움이 될 것이다.

앞에서 소개한 스마트 시대의 패러다임중 ‘빅데이터’ 개념이 도움이 될 수 있다. 일반적으로 특정 이러닝 클래스의 콘텐츠와 학습 운영 정책을 매학기 일정하게 유지가 된다. 다른 교과목 간에는 콘텐츠 내용의 분량이나 난이도가 상이하고, 담당 교수자의 강의 운영 정책도 다르다. 하지만 동일 교수자의 동일 교과목은 강의 운영 패턴이 대부분 유지된다. 따라서 학습자들의 행동 양상도 유사할 것이다. 이에 착안하여, Data based Getting Position Method (데이터에 기반한 위치 파악 방법), DPM을 제안한다. DPM 방법은, 강의 운영 패턴이 유사한 교과목의 수강생 데이터를 기반으로 현재 수강 진행중인 수강생의 Position을 파악할 수 있게 해준다.

작년의 특정 교수의 특정 교과목에 대한 수강생 데이터를 분석하여, 금년 동일 교과목을 수강하고 있는 학생에게 유용한 정보를 줄 수 있다. 예를 들어, 학기 중간 즈음 수강생이 성취한 지표들, 예를 들어 출결상황이나 과제점수, 중간고사 점수 따위가 있다. 이때 작년 수강생들 중 지금의 본인과 유사한 성취를 이룬 수강생의 데이터를 파악한다. 이 정보를 통해서, 수강생이 원하는 최종학점이 A 학점이라면 앞으로 어떤 부분에서 어느 정도의 성취를 해야 하는지를 알게 된다. 학기 중간까지 부족했던 부분을 알게 되어 보완할 수 있

고, 잘하고 있던 부분을 알게 되어 학습 의욕도가 높아질 수 있다.

교수자의 입장에서도 DPM은 유용한 수단이 될 수 있다. 기존의 클래스에서는, 학기 중간 과제 점수나 시험 점수 등을 단순 공시하는데 머물렀다면, 앞으로는 학습자 별로 보완점을 제시해 줄 수 있고 잘한 부분에 있어 칭찬을 해 줄 수 있게 된다.

DPM은 이렇듯 학습자와 교수자 모두에게 유용한 정보를 제공해 줄 수 있고, 미래지향적인 행동을 취할 수 있게 해준다. 이는 궁극적으로 학습 운영의 정성적 질을 높이는데 도움을 줄 것이다.

또한 이 방법은 2-3에서 다루었던 기존 이러닝 운영 모델의 한계점을 보완하는 방법이다. 교수자의 입장에서는 수백명의 학생들의 학습 정도와 현황을 유용하게 파악할 수 있게 되어 학습 운영 정책이나 강의 방법에 변화를 줄 수 있게 된다. 학습자 입장에서는 본인의 상대적인 위치를 가늠할 수 있게 되어 시간 낭비를 줄이고 본인의 핵심역량에 집중할 수 있게 해준다.

IV. 모형 설계 및 실험 결과

4.1. 대상 및 클래스 운영 정책

본 연구 실험의 대상은 인천의 A대학의 2014년 봄 학기에 개설된 교양 온라인 강의 I클래스의 수강자를 대상으로 하였다. 대상은 학습 중도 포기자들을 제외하고 472명의 수강생을 대상으로 하였다. 학습 운영 정책은 크게 7개 요인으로 분류되어 최종성적이 부여되었다.

7개 요인은 첫째 ‘자율과제’로, 과제 참여에 자율성을 부여하여 대상 수강생들의 참여도를 파악하기 위한 과제로 학습자들의 학습 적극성을 알 수 있는 요인이다. 둘째는 ‘필수과제’로, 일정 주제에 대한 리포트를 수행하는 과제였고 과제 점수를 4단계로 나누어 채점하였다. 세 번째로는 필수과제의 ‘과제완료시점’ 요인이다. 일반적으로 과제 제출 기한에 맞추어 내는 학생들이 있는 반면 기한보다 일찍 제출하는 학생과 늦게 제출하는 학생들로 나뉜다. 이 요인도 학생들의 학습 준비도를 파악할 수 있는 요인이다. 네 번째로는 ‘질문여부’ 요인이다. 1개 학기동안 온라인 클래스 운영 플

랫폼에는 게시판 형식으로 질의응답을 할 수 있는 코너가 마련되어 있다. 이 채널을 통해 학생들과 교수자가 질의응답 형식의 소통을 한다. 따라서 적극적으로 질문을 하려는 학생들은 학습 참여에 열의가 있는 것으로 볼 수 있기에 요인으로 정하였다. 다섯 번째는 ‘공지사항 확인 수’ 요인이다. 학기 중 온라인 클래스 게시판을 통해 학습정보나 과제 가이드 라인의 공지사항을 공개하고 있다. 학습자가 이러한 공지사항을 꼼꼼히 확인하는지 여부 역시 학습자의 참여도를 파악할 수 있는 요인이다. 여섯째로 ‘출석률’ 요인이다. 출석률이나 결석 수 등이 학습자의 학습자세를 가늠할 수 있는 요인이라고 판단하였다. 마지막으로 ‘기말고사 점수’ 요인이다. 클래스 운영시 가장 비중이 있는 요인으로, 학기말 최종적으로 서면 객관식 시험을 통해 학습도를 평가하는 요인이다.

학습자들의 7가지 요인들을 점수화 하여 최종 점수를 책정하고, 이를 A, B, C 3단계로 평가하여 클래스의 수강생 평점을 부여하였다.

4.2. 실험 데이터로의 변환

학기 중 취합된 472명 수강생들의 7가지 요인 데이터를 확보한 후, 객관적인 분석을 위하여 통계를 활용하기로 하였다. 통계분석에 앞서 7가지 요인의 실험 데이터로써 정리 작업이 필요하였다.

‘자율과제’ 요인은 과제 참여수의 평균을 내어 평균 이상으로 참여한 학습자, 평균 이하로 참여한 학습자 두 분류로 나누었다. ‘필수과제’ 요인은 A, B, C 3단계로 과제를 평가하고, 미제출 자는 F를 부여하여 4단계로 분류하였다. ‘과제완료시점’ 요인은 과제 제출 기한 이전에 제출한 학습자, 제출 기한 당일에 제출한 학습자, 기한을 넘겨서 제출한 학습자 3가지로 분류하였다. ‘질문여부’는 질문한 학습자와 질문을 하지 않은 학습자로 분류하였다. ‘공지 확인 수’는 전체 학습자의 공지사항 확인 횟수를 종합하여 평균 확인 횟수 이상, 이하 두 가지로 분류하였다. ‘출석률’ 요인은 전체 출석 횟수 중 결석수가 몇 회나 있었는지를 반영하였다. 마지막 요인인 ‘기말고사 점수’ 요인은 객관식 40문항의 시험을 50점 만점으로 환산하여 점수를 반영하였다.

최종 성적인 A, B, C 3단계 평점을 타겟으로 지정하고 이 변수의 이름을 ‘target_score’로 생성하였다. 7

가지 요인들과 'target_score'의 관련성을 통계 분석 하였다.

4.3. 가설 설정

실험결과를 분석하기 이전에 두 가지 가설을 설정하였다.

첫째, 온라인 클래스 운영에 있어서 학기 중 평가 요인들이 최종 성적인 target_score와 유의미한 연관이 있을 것이다. 일반적으로 비대면적인 특성의 온라인 클래스에서는 학습자들의 정성적인 학습 태도나 참여도, 적극성을 파악하기가 어렵다. 하지만 여러 가지 요인들의 참여 패턴을 빅데이터를 통해 분석하게 되면 수 백명 대상자들 중 비슷한 패턴을 가지는 학습자들이 분류가 될 것이고, 이러한 점들이 최종 성적과도 대입이 될 것이라고 예상하였다.

둘째, 유사한 운영 방식을 가진 다른 클래스에도 분석결과가 일반적으로 적용될 것이라고 예상하였다. 상이한 교과목은 학습 콘텐츠도 다르고, 수강생과 교수자도 다르다. 하지만 학습자들의 학습 패턴은 서로 다른 교과목들 사이에서도 일정한 패턴을 유지할 것이라고 보았고, 특히 학습 운영 정책이 같거나 유사한 교과목 사이에서는 같은 분석 결과를 보일 것이라고 판단하였다. 이를 확인하기 위하여 I클래스의 분석 결과를 동일 교수자, 동일 운영 정책으로 운영되었던 2014년 1학기 I 대학의 온라인 교양 교과목 M클래스에 적용하기로 하였다.

4.4. 모형 설계

첫단계로 7가지 요인 중 target_score와 연관이 있는 요인을 파악하기 위하여 빈도분석을 수행하였다.

<표1>은 학점(target_score)에 영향력을 줄 것 같은 요인 중 범주형 자료에 대한 빈도분석 결과이다. 빈도 분석에서 사용한 요인은 일반과제(homework), 필수과제(complete), 질문여부(ask_q), 그리고 공지확인 수(notice)가 있다. 위 표의 결과를 보면 일반과제에서 A을 받은 학생들은 총 21명중 19명(90.48%)이 A학점을 받은 것을 알 수 있었고, 미제출 한 학생은 모두 C학점을 받은 것을 볼 수 있다. 이를 볼 때, 일반과제가 학점을 구분 할 수 있는 기준이 될 수 있는 요인이라는 것을 예상해 볼 수 있다.

표 1. 학점과 범주형 요인들에 대한 빈도분석
Table. 1 Frequency analysis of target score and factors

| | | Target_score(학점) | | | |
|--------------------|----------------|------------------|-------|-------|-----|
| | | 1 (A) | 2 (B) | 3 (C) | 합계 |
| homework (일반과제) | 1 (A) | 19 | 1 | 1 | 21 |
| | 2 (B) | 25 | 10 | 3 | 38 |
| | 3 (C) | 95 | 171 | 143 | 409 |
| | 4 (F : 미제출) | 0 | 0 | 4 | 4 |
| complete (필수과제) | 0 (당일 제출) | 54 | 76 | 55 | 185 |
| | 1 (일찍 제출) | 81 | 86 | 37 | 204 |
| | 2 (늦게 제출) | 4 | 20 | 59 | 83 |
| ask_q (질문여부) | 1 (질문 함) | 54 | 55 | 42 | 151 |
| | 2 (질문 안함) | 85 | 127 | 109 | 321 |
| notice (공지확인 수) | 1 (평균 이상) | 105 | 110 | 61 | 276 |
| | 2 (평균 이하) | 34 | 72 | 90 | 196 |
| 합계 | | 139 | 182 | 151 | 472 |

좀 더 정확한 차이를 확인하기 위해 교차분석을 통해 각 요인과 학점이 서로 독립인가를 확인 한 결과는 아래 <표2>와 같다.

표 2. 학점과 각 요인들에 대한 분석 결과
Table. 2 Analysis result of target score and factor

| 요인 | Test | P-value |
|----------|---------------------|-----------|
| homework | Fisher's exact test | 5.589E-15 |
| complete | Chisquare test | <.0001 |
| ask_q | Chisquare test | 0.1066 |
| notice | Chisquare test | <.0001 |

검정 결과를 볼 때, 학점에 영향을 줄 것이라고 생각했던 요인들이 유의수준 0.05하에서 모두 학점과 독립이 아님을 알 수 있었다. 즉, 학점과 관련이 있는 요인이라고 생각 할 수 있기 때문에 위의 네 가지 요인들을 Modeling에 사용하였다.

위 네 가지 요인 외에 연속형 자료에 속하는 결석 횟수(absent)와 기말 점수(final)도 학점에 따라 차이가 있는지를 확인하기 위해 분산분석(ANOVA)를 실시

하였다.

표 3. 기말점수(final)와 학점(target_score)에 대한 분산분석 결과

Table. 3 Statistical dispersion analysis of final and target score

| Source | DF | SS | MS | F value | P-value |
|--------|-----|-------------|------------|---------|---------|
| Model | 2 | 14756.28055 | 7378.14028 | 424.75 | <.0001 |
| Error | 469 | 8146.81466 | 17.37061 | | |
| Total | 471 | 22903.09521 | | | |

<표3>은 기말점수(final)와 학점(target_score)에 대한 분산분석 결과이다. 위 결과를 보면 유의수준 0.05하에서 P-value가 0.0001보다 작은 것으로 보아 유의한 결과임을 알 수 있고, 이는 학점에 따른 기말점수는 동일하다는 가설을 기각 하므로, 적어도 하나는 다르다는 것을 알 수 있다. 그 차이를 더 정확하기 위한 분석으로 Tukey의 사후 분석을 사용하였고 그 결과는 아래 <표 4>에 나타난다.

표 4. 기말점수(final)와 학점(target_score)에 대한 사후분석

Table. 4 Postmortem result of final and target score

| Comparisons significant at the 0.05 level are indicated by ***. | | | | |
|---|--------------------------|------------------------------------|----------|-----|
| target_score Comparison | Difference Between Means | Simultaneous 95% Confidence Limits | | |
| 1 - 2 | 6.9917 | 5.8879 | 8.0955 | *** |
| 1 - 3 | 14.2684 | 13.1166 | 15.4203 | *** |
| 2 - 1 | -6.9917 | -8.0955 | -5.8879 | *** |
| 2 - 3 | 7.2767 | 6.1980 | 8.3554 | *** |
| 3 - 1 | -14.2684 | -15.4203 | -13.1166 | *** |
| 3 - 2 | -7.2767 | -8.3554 | -6.1980 | *** |

target_score Comparison은 각 학점별의 평균의 차이를 의미한다. 예를들어 1 ? 2의 경우, 1에 해당하는 A학점과 2에 해당하는 B학점에서의 기말고사의 평균점수의 차이를 나타내고 그 차이가 6.9917이라는 것을 알 수 있다. 이를 볼 때, 모든 경우의 수가 유의하다는 것을 알 수 있고, 이는 A학점, B학점, 그리고 C학점의 기말점수의 평균이 모두 유의하게 차이가 있다고 생각 할 수 있다.

다음으로는 결석 횟수(absent)에 대한 분산분석 결과가 <표5>에 나타나있다.

표 5. 결석 횟수(absent)와 학점(target_score)에 대한 분산분석 결과

Table. 5 Statistical dispersion analysis result of absent and target score

| Source | DF | SS | MS | F value | P-value |
|--------|-----|-------------|------------|---------|---------|
| Model | 2 | 382.136785 | 191.068392 | 38.85 | <.0001 |
| Error | 469 | 2306.431012 | 4.917763 | | |
| Total | 471 | 2688.567797 | | | |

결석 횟수(absent)와 학점(target_score)에 대한 분산분석 결과에서도 유의수준 0.05하에서 P-value값이 0.0001보다 작으므로 학점 별 결석 횟수의 평균이 차이가 없다는 귀무가설을 기각 할 수 있다. 그러므로 적어도 하나는 평균이 다르다는 것을 알 수 있고, 이를 더 자세히 확인하기 위한 Tukey의 사후검정 결과가 아래 <표6>에 정리되어있다.

표 6. 결석 횟수(absent)와 학점(target_score)의 사후분석

Table. 6 Postmortem result of absent and target score

| Comparisons significant at the 0.05 level are indicated by ***. | | | | |
|---|--------------------------|------------------------------------|---------|-----|
| target_score Comparison | Difference Between Means | Simultaneous 95% Confidence Limits | | |
| 3 - 2 | 1.5756 | 1.0017 | 2.1496 | *** |
| 3 - 1 | 2.2056 | 1.5927 | 2.8185 | *** |
| 2 - 3 | -1.5756 | -2.1496 | -1.0017 | *** |
| 2 - 1 | 0.6300 | 0.0426 | 1.2173 | *** |
| 1 - 3 | -2.2056 | -2.8185 | -1.5927 | *** |
| 1 - 2 | -0.6300 | -1.2173 | -0.0426 | *** |

결석 횟수(absent)와 학점(target_score)도 기말점수와 마찬가지로 모든 경우에서 유의한 차이가 있다고 볼 수 있기 때문에 각 학점 별로 결석 횟수의 평균의 차이가 있다고 생각 할 수 있다.

위 결과를 바탕으로 학점을 판별할 수 있는 Model을 생성하기 위하여 의사결정 나무(Decision Tree)방법을 선택하였고, 신경망(Neural Network)은 해석이 어렵기 때문에 분석의 목적과 맞지 않을 것이라고 판단하여 분

석에서 제외하였다.

4.5. 통계분석결과

분석 결과는 <그림 3>과 같은 형태가 생성되었고, 최종 선택된 요인은 중요한 변수 순서로 기말점수(final), 결석 횟수(absent), 필수과제(complete), 일반과제(homework)이다.

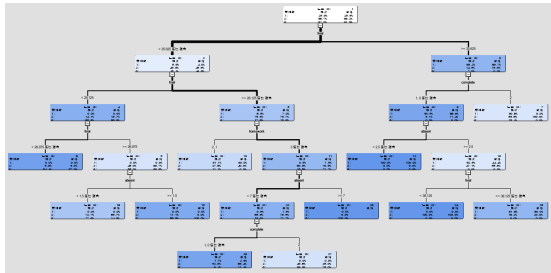


그림 3. 의사결정나무
Fig. 3 Decision tree

위 분석은 70%의 데이터로 모델링을 하고 나머지 30%의 데이터로 Test를 하여 확인 하였고, 아래 <표7>은 30%의 데이터(Validation set)으로 test를 한 결과이다.

표 7. 30%의 데이터(Validation set)에서의 결과
Table. 7 Result of 30% Validation set

| 타겟 | 결과 | 타겟 백분율 | 결과 백분율 | 빈도수 | 총 백분율 |
|--------|----|----------|---------|-----|---------|
| 1 | 1 | 93.3333 | 97.6744 | 42 | 29.1667 |
| 2 | 1 | 6.6667 | 5.4545 | 3 | 2.0833 |
| 1 | 2 | 1.7857 | 2.3256 | 1 | 0.6944 |
| 2 | 2 | 87.5000 | 89.0909 | 49 | 34.0278 |
| 3 | 2 | 10.7143 | 13.0435 | 6 | 4.1667 |
| 2 | 3 | 6.9767 | 5.4545 | 3 | 2.0833 |
| 3 | 3 | 93.0233 | 86.9565 | 40 | 27.7778 |
| 오분류 비율 | | 0.090278 | | | |

<표7>에서의 결과에서 오분류 비율을 확인 해 보면 0.090278로 약 9%를 제외하고 나머지 약 91%에 해당하는 값은 잘 예측되었다는 것을 알 수 있다.

V. 결과 정리 및 응용

이러닝 클래스 운영 정책 요인들 중 최종 학점과 유의한 연관이 있는 요인들을 분석하였다. 학기 중 학습자들의 과제 제출기한 준수 여부, 과제 점수, 질문 횟수, 공지사항 확인 수 등이 최종 학점과 관련이 있었다. 또한 기말고사 점수나 출석 빈도 역시 최종 학점인 A, B, C학점의 분류에 유의하게 영향이 있는 점을 확인하였다.

학기 중 학습 요인의 분석을 통한 최종 학업 성취도 예상 모델을 생성하기 위하여 의사결정나무(Decision Tree) 방법을 선택하였다. 의사결정나무 방법으로 요인들을 분류하기에 앞서, 최종 학점 구분을 위한 중요 변수로는 기말고사점수, 결석횟수, 과제점수 순이었다. 이 분석 결과를 토대로 의사결정나무 모델을 완성한 후, Row data를 모델에 재 대입시켜 검토한 결과 오분류 비율이 약 9% 정도로 산출되어 나머지 약 91%에 해당하는 값은 잘 예측된다는 것을 알 수 있었다. 따라서 4.3에서 설정하였던 두 가지 가설 중 첫 번째 가설인 학기 중 평가 요인과 최종 성적 간의 유의미한 연관성이 있음을 확인 하여 가설을 채택할 수 있다.

본 모델의 일반성을 확보하기 위한 두 번째 가설을 확인하기 위해서, 같은 교수자의 다른 교과목 클래스 M 과목에 실험에서 만들어진 의사결정나무 모델을 대입시켜 보았다. 두 교과목은 학습 운영 정책이 동일하며, 둘 과목 모두 이러닝 교양 교과목으로 학습 테마만 다른 교과목이다. 따라서 유사한 운영 정책의 클래스에서는 요인들의 패턴도 유사할 것이라고 보았다. M교과목에 모델을 대입시킨 결과, 오분류 비율이 약 11.38%로 나타나 전체중 약 88.62%에 해당하는 값들은 잘 예측이 되었음을 확인할 수 있다. 따라서 두 번째 가설 역시 채택할 수 있다.

VI. 결론 및 과제

본 논문은 온라인 학습 환경과 산업이 활발하게 성장하고 있는 사회에서 구체적인 모델 제안과 실증을 통한 연구를 진행하였다. 각 기관이나 학교 등에서 쓰이고 있는 온라인 학습 운영 모델이 그 경제성에만 집중되어 운영되는 환경에서 어떻게 하면 더 적절하게 온라인의

장점을 활용할 수 있는가에 초점을 두었다.

빅데이터라는 개념을 온라인 학습에 접목시켜, 교수자와 학습자가 학습 과정에 있어 객관적이며 실용적인 지표를 얻을 수 있는 모델을 생성하였다. 학습 과정 중간에 학습 요인들의 상대적인 위치를 확인할 수 있도록 하여 부족한 점과 잘하는 점을 파악할 수 있고, 이를 통해 최종 학점에 대한 예측이 어느 정도 가능하도록 하였다. 이를 활용하면 학습자의 학습의지 상승에 도움을 줄 수 있는 동시에, 교수자는 학습자들의 객관적인 지표와 개개인적인 특성을 쉽게 파악할 수 있다는 장점이 있다.

본 연구에 몇 가지 한계점을 꼽는다면, 첫째로 기존에 운영되는 수많은 온라인 클래스들의 운영모델이 상당히 다양하다는 점이다. 학습내용의 특성이나 학습자들의 수준, 교수자의 운영 정책 등이 모두 제각각이며 운영 요인에 대한 가중치도 상이할 수 있다. 따라서 다양한 운영 정책을 갖고 있는 클래스에도 더욱 일반적으로 적용할 수 있는 아이디어에 대한 연구가 필요할 것으로 보인다.

둘째로, 오분류의 경우를 어떻게 다룰 것인가에 대한 문제이다. 실험 결과에서는 약 90% 정도로 최종 학습성취도를 예측하였으나, 학교 등 공적인 성과를 도출해야 하는 환경에서는 약 10% 내외의 오분류 경우를 적절하게 처리할 수 있는 방안이 필요하다.

본 연구는 학습자들의 행위양상을 교수자의 학습의도와는 다를 수 있다는 아이디어에서 출발하였다. 전통적인 교실에서는 교수자의 강의에서의 의도가 학습자에게 잘 전달되고 있는지를 파악하기 어려웠다. 학습자 역시 그 의도를 잘 전달받고 있는지에 대한 확신은 없었다. 이 간극을 객관적인 지표로 구체화 하는 작업을 수행하였고, 그 과정에서 스마트 시대의 도구들을 활용하였다. 이를 통해 학습자와 교수자가 얻는 의미성에 대한 연구를 지속할 예정이고, 온라인 클래스 내에서

미래 예측 가능성을 조금 더 높이기 위한 실험을 계속 진행 할 것이다. 궁극적으로 무한 소통의 시대에 조금 더 사람들 간의 커뮤니케이션을 원활하게 하기 위한 스마트 러닝을 구현할 것이다.

감사의 글

본 연구는 2015년도 인하대학교 지원에 의하여 이루어진 연구로서, 관계부처에 감사 드립니다.

REFERENCES

- [1] J. W. Yoon, "the introduction of Smart Media era and develop", *Korea Society Broadcast Engineers Magazine*, vol. 18, no. 1, pp. 10-22, Jan. 2013.
- [2] Lee, Jong-Man, "What Drives a Successful e-Learning: Focusing on the Critical Factors Influencing e-Learning Satisfaction", *Korea Journal of Business Administration*, vol.24, no. 4, pp. 2245-2257, Aug. 2011.
- [3] Doo Young Lee, "A Study on the Improvement of Online Education Model", M.S. dissertation, INHA University, 2014.
- [4] Jae Chun Park, *Internet Story*, 1th ed. INHA Univ, 2013.
- [5] Tim O'REILLY, "What Is Web 2.0: Design Patterns and Business Models for the Next Generation of Software", *COMMUNICATIONS & STRATEGIES*, no. 65, 1st quarter 2007, p. 17.
- [6] Akshay Java, Tim Finin, Xiaodan Song, Belle Tseng, "Why We Twitter: Understanding Microblogging Usage and Communities", *9th WEBKDD and 1st SNA-KDD Workshop '07*, August 12, 2007, San Jose, California, USA.



박재천(Jae-Chun Park)

서울대학교 공과대학 응용수학과 공학사
 Georgia Institute of Technology, O.R. 공학석사
 University of Hawaii at Manoa, Inform. & Commun. Economics 경제학박사
 현 인하대학교 교수
 ※ 관심분야 : 정보통신정책, IT산업정책, 이러닝



이두영(Doo-Young Lee)

인하대학교 기계공학과 공학사
인하대학교 일반대학원 IT미디어융합전공 공학석사
현 인하대학교 일반대학원 IT미디어융합전공 박사과정
※관심분야 : 정보통신정책, IT산업정책, 이러닝



국성희(Sung-Hee Kuk)

인하대학교 통계학과 학사
인하대학교 일반대학원 통계학과 석사
※관심분야 : 데이터마이닝, 이러닝