

맞춤형 학습코스 추천 모델의 효과분석 방안

한지원† · 임희석†

† 고려대학교 컴퓨터학과

Analysis of the effectiveness of the Recommendation Model for the Customized Learning Course

Ji-won Han† · Heui-seok Lim†

† Dept. of Computer Science and Engineering, Korea University

요 약

본 논문은 사용자 수준에 적합한 맞춤형 학습코스를 추천하여 학습효과를 향상시킬 수 있는 추천모델을 개발하고, 효과분석을 위한 방안을 제시한다. 학습자 개인의 학습수준이나 학습내용 등에 따라 적합한 학습주제를 선정하여 제공하는 것은 중요하나, 일반적인 추천은 전문가 그룹을 활용한 사람중심의 추천으로 시간이 오래 걸리는 등 자원의 비효율적 한계점[1]을 가지고 있다. 이를 극복하기 위해, TF-IDF를 이용해 단어별 가중치를 계산하여 고빈도 단어를 추출하여 벡터 공간에 배치시키고, Cosine Similarity 기법을 이용해 벡터간의 유사도를 측정하였다. 학습자 프로파일을 분석하고, 학습스킬간의 연관성을 고려하여 맞춤형 학습코스를 추천하기 위해, 워드 임베딩 기법을 적용하였고, 이를 위해 오픈소스 Gensim[2]을 이용하였다. 맞춤형 학습코스 추천 모델의 효과를 분석하기 위한 실험을 설계하고 평가 문항지를 개발하였다.

1. 서 론

정보통신기술(ICT, Information & Communication Technology)의 발달과 ICT를 이용한 인프라의 확장으로 사람들은 언제 어디서나 정보를 생산하고, 공유할 수 있게 되었으며 다양한 방식으로 의사소통을 진행하는 것이 가능하게 되었다. 스마트 디바이스의 등장으로 비형식 학습 활동을 할 수 있는 환경이 구축되면서 학습자간 상호작용이 활발해졌고, 온라인 커뮤니티를 이용한 지식공유 활동이 활성화 되었다. 교육 분야에서도 ICT 기술이 적용되어 새로운 패러다임으로 변화하고 있다. 전통적인 교육 방식에서는 교육 과정을 편성하고 교사와 학생간의 면대면 학습방식을 제공하는 체계적인 형식 학습(Formal Learning)이 주를 이루고 있다. 그러나 ICT 기술의 발전으로 시·공간 제약의 한계가 극복되고 정규교육 과정뿐만 아니라 평생 교육(Lifelong Education)과 같이 다양한 연령대와 직업을 가진 사람들이 교육에 참여하고 있다. 이에 따라 형식적 교육 방식에서 ICT 기술을 활용한 정보 검색, 지식 구성, 협력적 상호작용 등을 통하여 경험을 확대하고 학습이 이루어지는 비형식 학습(Non-formal Learning)으로 교육의 흐름이 변화하고 있다. 다양하고 많은 정보들과 이를 검색할 수 있는 정보검색 시스템은 다양한 방법으로 학습자의 지식을 확장하는데 도움을 주지만 기본적으로 학습과 경험은 개인의 편차가 심하고 학습효과도 분절적으로 발생하기 쉽기 때문에 단순히

의미미한 정보 전달 및 학습 경험이 학습 효과 향상에 영향을 미치지 못한다. 또한, 학습 분야에 대한 전문성이 없는 학습자들에게는 어떠한 학습 정보가 의미미한지 분별하기 어려우며 교수자가 없이 자기주도 학습을 진행해야 하는 비형식 학습 환경의 한계로 인하여 학습에서 무엇을 학습해야 하는가에 대한 것을 스스로 결정하는데 어려움을 느끼고 있다.

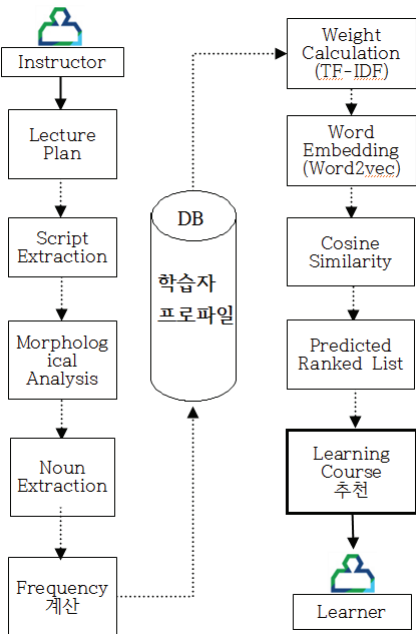
이를 극복하기 위해서 추천 시스템이 교육환경에 적용되었다. 추천 시스템은 사용자의 과거 행동 데이터나 다른 데이터를 바탕으로 사용자에게 필요한 정보나 제품을 사용자의 성향에 맞춰 추천함으로써, 사용자들은 자신이 원하는 품목을 쉽게 획득할 수 있도록 돕는 시스템이다. 이러한 추천 시스템은 개인의 성향에 따라 효과가 다르게 나타나는 교육 환경에서도 필요한 부분이며 현재 다양한 방법으로 교육 분야에 응용되고 있다. 특히, 선택에 대한 어려움과 시행착오를 줄여 학습효과를 증진시킬 수 있도록 학습자 수준에 적합한 학습코스를 제공해 줄 수 있다는 점에서 추천 시스템은 비형식 학습 환경에서 매우 유용하다. 이는 학습자 개인의 학습수준이나 학습내용 등에 따라 적합한 학습주제를 선정하여 제공해 줌으로써, 학습자 스스로가 자신의 학습과정을 계획하고 구성하는 것이 가능해지기 때문이다.

이에, 본 연구는 사용자 수준에 적합한 학습코스를 추천하여 학습효과를 향상시킬 수 있는 추천시스템을 제안하고자 한다. 인간중심의 SSDACUM 모델을 개발

하여 추천했던 기존연구[1]의 한계점을 극복하여 맞춤형 추천모델을 개발하여 학습코스를 추천한다. 향후, 개발한 추천시스템의 효과를 분석하고 검증하기 위해 만족도조사를 실시하고자 한다.

2. 맞춤형 학습코스 추천모델

본 연구는 인간중심의 전문가 그룹으로 구성되어 학습코스를 추천하는 비효율성을 극복하기 위해, 기계학습 중심의 추천모델을 개발하여 제안한다[1]. 이를 위해 TF-IDF (Term Frequency-Inverse Document Frequency)를 이용해 단어별 가중치를 계산하여 고빈도 단어를 추출하여 벡터 공간에 배치시킨다. TF-IDF는 정보검색과 텍스트 마이닝에서 이용하는 가중치로, 여러 문서로 이루어진 문서군이 있을 때 어떤 단어가 특정 문서 내에서 얼마나 중요한 것인지를 나타내는 통계적 수치값을 알 수 있다. TF-IDF를 통해 추출된 단어를 벡터 공간에 배치시켜, Cosine Similarity 기법을 이용해 벡터간의 유사도를 측정한다. 학습자 수준 및 학습 콘텐츠의 구성요소 간 연관성을 고려하여 학습코스를 추천하기 위해, 워드 임베딩 기법(Word Embedding) 기법을 적용한다. 이는 신경망언어모델로부터 도출된 기술로 유사한 단어를 벡터 공간상에 가깝게 배치하여 어휘 의미를 표현할 수 있는 기술이다. 기계학습 방법을 활용한 예측된 학습코스를 학습자에게 추천하는 것이 가능하다. 제안하는 모델의 전체 구조도는 [그림 1]과 같다.



[그림 1] 기계학습 중심 학습코스 추천모델

2.1 고빈도 단어 추출

본 모델은 학습자 프로파일에 기반한 기계학습 중심의 학습코스를 예측하기 위하여, 우선 교수자가 강의안을 업로드한다. 업로드된 강의안의 내용중에서 스크립트를 추출하여 형태소 분석기를 통해 각 단어 별로 형태소가 태그된다. 형태소가 태그된 단어들 중에서 명사 단어만 추출하고 그 단어의 빈도수를 계산하여 DB에 명사 단어와 빈도수를 저장한다.

시스템은 DB로부터 단어와 빈도값을 호출하고 각 단어 별 가중치를 계산하여 고빈도 단어들을 추출한다. 즉, 문서에서 많이 나타나는 명사가 그 문서를 대표하는 중요한 명사이므로 한 문서에 출현한 명사의 빈도수를 문서의 전체 길이로 정규화한 TF(Term Frequency)를 이용한다. 또한, 명사의 문서 간 가중치는 전체 문서집합 가운데 해당 명사가 나타난 문서가 어느 정도인지를 수치화한 IDF(Inverse Document Frequency)를 이용한다. 이는 특정 문서에만 나타나는 명사는 가중치는 높이고, 대부분의 문서에 나타나는 범용 명사는 가중치를 낮추는 역할을 한다. 위의 2가지 가중치를 함께 고려하기 위해 각 문서에 속한 명사의 최종 가중치는 TF x IDF로 결정되며, 이 값이 높은 명사들을 대표 명사목록으로 선택하게 된다. 가중치(TF-IDF) 계산식은 아래와 같다.

$$W_{ij} = tf_{ij} \times \left\{ \log\left(\frac{N}{n}\right) + 1 \right\}$$

W_{ij} : j번째 강의에서 i번째 단어의 가중치

tf_{ij} : j번째 강의에서 i번째 단어의 빈도

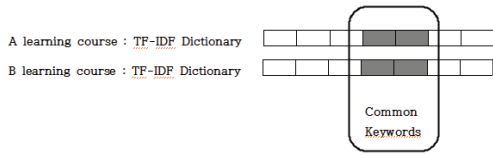
N : 학습자가 현재까지 완료한 강의의 수

n : N 에서 W_{ij} 단어를 포함하고 있는 강의수

TF-IDF가 높은 단어들은 문서의 Identity를 더 높게 반영한다고 할 수 있고, 이 값들이 유사하다면 유사한 문서라고 생각할 수 있다. 벡터간의 유사도를 측정하는 여러 방법이 있지만, 그 중 Cosine Similarity 기법이 있다. 계산식은 아래와 같다.

$$\cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i \times B_i}{\sqrt{\sum_{i=1}^n (A_i)^2} \times \sqrt{\sum_{i=1}^n (B_i)^2}}$$

비교할 문서들의 TF-IDF 값을 담은 Dictionary 자체를 벡터로 하여 문서간의 유사도를 구할 수 있다. 즉 분자에는 공통된 부분의 Dot 곱이, 분모에는 각 Dictionary 벡터들의 크기가 들어가며, 시스템 구성과 코드로 옮겨보면 아래 [그림 2]와 같다.



```
def cosine similarity(vector a, vector b, vector dot):
    a vector magnitude=0
    for elem in vector a:
        a vector magnitude += math.pow(elem, 2)
    b vector magnitude=0
    for elem in vector b:
        b vector magnitude += math.pow(elem, 2)

    magnitude=math.sqrt(a vector
        magnitude)*math.sqrt(b vector magnitude)
    dot product=0
    for d in vector dot:
        dot product += d

    return dot product/magnitude
```

[그림 2] 문서 간 유사도 계산 코딩

3. 교육효과 분석

3.1 실험설계

기계학습 중심의 학습코스를 예측하여 추천하는 본 연구의 효과를 분석하기 위해, 실험집단(G1)과 비교집단(G2)을 선정하여 실험하였다. 두 집단 모두 경기도 소재의 교육기관의 동일한 학습코스를 수료한 교육생으로 구성하였다. 학습수준을 정량평가 70점, 정성평가 30점으로 총합 100점을 기준으로 하여 취득한 점수의 평균에 따라 프로파일을 생성했다. 정량평가는 출석률 30점, 과제물 제출 20점, Final TEST 20점, 정성평가는 각 과제별 담당강사가 학습자의 교육 참여도 30점을 합산하여 평가점수를 도출한다. 학습기간 동안 교육 참여율 및 과제 미제출 등으로 교육을 수료하지 못한 교육생은 0점으로 처리한다. 그 결과, 70점미만인 학습자가 G1 32%, G2 37%, 71점~89점 학습자 G1 51%, G2 47%, 90점 이상인 학습자 G1 16%, G2 15% 나타났다. 두 집단의 학습수준 차이가 크지 않은 것으로 보아, 동일한 집단임이 검증되었다.

동일한 두 집단을 대상으로 전문가 그룹을 활용한

인간중심의 학습코스를 추천하고, 본 연구에서 제안하는 기계학습 중심의 학습코스를 추천한다. 추천받은 학습코스에 대해서 만족도조사를 통하여 효과성을 검증하고자 한다. 이를 정리하면 아래의 <표 1>과 같다.

<표 1> 실험설계

Group	Pre-test	실험처치	Post-test
G1	T1	X1	T2-1
G2		X2	T2-2

- G1 : 실험 Group
- G2 : 비교 Group
- T1 : 학습자 능력수준 평가
- T2-1, T2-2 : 교육효과성 평가
- X1 : 맞춤형 학습코스 추천
- X2 : 학습코스 추천[1]

3.2 학습자 프로파일

두 집단 모두 경기도 소재 교육기관에서 동일한 IT 전문교육을 수료한 교육생으로 구성했다. 교수자는 교실 당 1명씩 20명의 학습자를 통제한다. 학습자는 교육센터의 특성상 미취업여성으로 평균 연령 39세이며, IT를 전공한 교육생이 61%, 대학이상의 교육을 받은 학습자가 92%, 추천시스템을 사용 해본 경험이 있는 학습자는 47%로 나타났다.

<표 2> 학습자 프로파일

Dimension		Freq.(%)	
		G1	G2
Gender	Female	80(100)	180(100)
	Male	0(0.0)	0(0.0)
Age	20~29	8(10)	11(6.1)
	30~39	43(53.8)	110(61.1)
	40~49	26(32.4)	57(31.6)
	>=50	3(3.8)	2(1.1)
ICT majors	Yes	49(61.3)	98(54.4)
	No	31(38.7)	82(45.6)
Educational Attainment	High school	6(7.5)	22(12.5)
	Junior college	9(11.3)	27(15.0)
	Over college	65(81.3)	131(72.5)
Recommender System Experience	Yes	38(47.5)	19(10.5)
	No	42(52.5)	161(89.5)
Learning Proficiency (Mean values)	<=70	26(32.5)	68(37.8)
	71~89	41(51.3)	85(47.3)
	90~100	13(16.3)	27(15.0)

- G1 : 실험 Group(맞춤형 학습코스 추천)
- G2 : 비교 Group

실험집단(G1)교실은 2016년 8월부터 12월까지 16주간(총 400시간) 매주 5회씩 1일 5시간, 80일 수업이 진행되었고, 한 차시 당 50분 수업, 10분간 휴식으로 진행되었다. 비교집단(G2)교실은 2015년 3월부터 7월까지 16주간(총 400시간) 매주 5회씩 1일 5시간, 80일 수업이 진행되었고, 한 차시 당 50분 수업, 10분간 휴식으로 진행되었다. 두 집단(G1, G2)의 교실 모두 원활한 IT교육 환경 기반 조성을 위하여 각 교실마다 최신 사양의 하드웨어와 소프트웨어를 설치하여 운영하였다. 실험 기간 동안 매 수업 마다 운영현황을 파악하여 매주 보고서를 작성하여 통합 월간 운영 보고서를 작성하였다. 보고서에는 작성자, 일시, 과정명, 교사명, 운영 내용, 요구사항 등이 포함되어 있다. 학습자 프로파일은 위의 <표 2>와 같다.

3.3 평가 문항지 개발

맞춤형 학습코스 추천에 대한 효과를 분석하고 검증하기 위하여 이미 많은 연구에서 사용하고 있으며 PISA technical report 에서도 중요하게 강조한 만족도 설문조사를 실시하였다. 설문지는 15문항의 질문과 측정 항목은 리커트 5점 척도를 통해 측정하였으며, 10명의 전문가가 참여한 Delphi 기법을 통해 문항타당도를 평가하여 설문지를 개발하였다. 개발된 설문지를 바탕으로 자기기입식 설문조사 방법을 실시하여 추천시스템의 효과를 분석하고 검증하고자 한다.

<표 3> 교육효과성 평가 문항지

구분	세부내용	EFA	CA
정확성	내게 필요한 추천결과를 제공한다.	0.802	0.910
	학습코스 선택에 도움이 된다.	0.791	
	내게 적절한 학습코스를 제공한다.	0.773	
새로움	학습코스에 대한 호기심을 만족시킨다.	0.796	0.756
	추천한 학습코스는 내게 친숙하다.	0.698	
	관심에 맞는 새로운 코스를 제공한다.	0.754	
만족	추천시스템에 매우 만족한다.	0.782	0.918
	내가 사용한 추천시스템 중 만족한다.	0.709	
	추천을 사용한 내 선택은 지혜로웠다.	0.755	
자기참조	추천결과가 나와 연관이 있다.	0.727	0.915
	추천페이지가 나에게 잘 설명해준다.	0.768	
	추천시스템이 나의 흥미를 고려하여 학습코스를 제공했다고 생각된다.	0.731	
유용성	개인의 학습에 매우 유용하다.	0.785	0.857
	학습능력을 향상시킬 것이라 생각된다.	0.710	
	학습 집중에 유용할 것이라 생각된다.	0.734	

실험 참가자를 대상으로 정확성, 새로움, 만족, 자기참조, 유용성에 대한 설문을 진행했으며, 설문 문항은 선행연구에 의해 제시된 것을 본 연구 내용에 맞게 수정하여 사용하였다. 측정된 요인들에 대한 타당성을 측정하기 위해 탐색적 요인분석(EFA, Exploratory Factor Analysis)를 실시하였으며, 설문 항목에 대한 신뢰도를 측정하였다. 설문항목이 신뢰성을 보장받기 위해서는 신뢰도 계수(CA, Cronbach's alpha) 값이 0.7 이상이어야 하는데, 모든 변수가 0.7보다 높은 값을 나타내어, 설문항목 모두 신뢰성을 인정 받았다. 세부내용은 위의 <표 3>과 같다.

4. 결론 및 논의

전문가 그룹을 활용한 인간중심의 예측 추천모델이 가지고 있는 시간이 오래 걸리는 비효율적 한계점을 극복하기 위해 기계학습 중심의 학습코스를 추천하는 모델을 제안하고, 교육효과를 분석하고자 했다.

추천하는 학습코스는 학습자 개개인의 선행 학습 정도와 수준에 따라 개개인의 학습동기 유발 방법 및 학습 이해도 향상을 고려한 수준별, 단계별 교수 방법과 학습자를 고려한 맞춤형 학습코스 추천을 해야 할 것이다. 이를 위해, 학습자 프로파일을 분석하고, 학습스킬 간의 연관성을 고려하여 추천하는 모델을 개발하여 제안하고자 했다. 또한, 제안하는 추천모델의 효과성을 검증하기 위한 실험을 설계하고 평가 문항지를 개발하였다. 향후, 본 연구에서 제안하는 맞춤형 학습코스 추천모델의 효과성을 분석하기 위해 개발된 문항지를 바탕으로 만족도 조사를 실시 하고자 한다. 조사결과를 바탕으로, 항목별 평균, 표준편차, 왜도 첨도값을 계산하여 정량적 분석을 실시하여 효과성을 검증하고자 한다.

참고 문헌

- [1] Han, J. W., Jo, J. C., Ji, H. S., & Lim, H. S. (2016). A collaborative recommender system for learning courses considering the relevance of a learner's learning skills. Cluster Computing, 19(4), 2273-2284.
- [2] Rehurek, R., Sojka, P.(2010). Software framework for topic modelling with large corpora. In Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks, 46-50.