

# 협업 필터링을 활용한 학사관리 시스템

박준형, 최예린, 이하연, 금종민, 김영중 송실대학교 소프트웨어학부  
dbapjh@soongsil.ac.kr, yeiyeilin@soongsil.ac.kr, youngjong@ssu.ac.kr

## Academic Management System Using Collaborative Filtering

Junhyung Park, Yelin Choi, Hayeon Lee, Jongmin Keum,  
Youngjong Kim  
\*School of Software, Soongsil University

### 요 약

협업 필터링은 사용자들간의 집단 지성에 기반한 추천 알고리즘이다. 다양한 시스템에 쓰이고 있는 협업 필터링과 그 장단점에 대해서 알아본다.

그리고 본 논문에서는 기존의 학사 관리 시스템에 협업 필터링을 적용한 새로운 학사관리 시스템을 제안한다.

### 1. 서론

학생들은 수강 신청을 하고 수업을 들으면서 자신의 역량을 키워나가고, 좋은 학점을 얻는 것을 목표로 한다. 하지만 그 과목의 수업을 듣기 전까지는 수업에 대한 정보를 알기가 쉽지 않다. 우리는 이전에 수업을 들었던 학생의 수, 학생들의 과목 드랍률, 학점 등을 활용하여 수강 예정인 학생들에게 도움이 되고자 한다. 다른 사람들의 후기에만 의존했던 기존의 방식을 탈피하여 빅데이터 분석을 통한 결과를 이용한 합리적인 선택을 돕는 시스템을 제안하는 것이 본 논문의 목적이다.

이용자가 원하는 정보를 먼저 탐색하여 제공할 수 있다는 점에서 정보 필터링 시스템은 최근 온라인 마케팅을 포함한 다양한 분야에서 활용되고 있는데, 본 논문에서는 그중 협업 필터링을 이용함으로써 더 나은 추천 정보를 제공하려 한다. 협업 필터링 알고리즘과 그가 가지는 문제점을 알아보고, 직접 학사관리 시스템에 적용해본다.

### 2. 본론

#### 2.1 협업 필터링

협업 필터링 알고리즘은 기존 사용자들의 아이템에 대한 평가기록을 바탕으로 다른 사용자들에게 아이템을 추천해주는 방법이다. 아이템에 대한 평가 기록은, 평가 행렬에 저장한다.

협업 필터링 알고리즘은 사용자 기반 알고리즘과 아이템 기반 알고리즘으로 나누어진다. [1]사용자 기반 알고리즘은 선호도가 유사한 사용자들을 근접 이웃으로 지정하고, 이 근접 이웃들의 아이템 평가 기록을

넣은 아이템 평가 행렬을 기반으로 사용자에게 아이템을 추천하는 기법이다. 아이템 기반 알고리즘은 사용자 기반 알고리즘과는 다르게 목표 사용자가 선호했던 아이템을 분석해 그 아이템과 유사한 다른 아이템을 추천하는 알고리즘이다. 협업 필터에는 초기 평가, 확장성, 희박성의 문제점이 있다.[2] 초기 평가란, 평가기록이 없는 상태를 말한다. 협업 필터링은 이미 존재하는 데이터에 의존해서 새로운 아이템을 추천하는데, 새로운 사용자나 평가된 적이 없는 새 아이템의 경우에는 데이터가 없으므로 이에 대해서는 추천하기 어렵다.

확장성 문제는 사용자와 아이템의 수가 늘어날수록 탐색 연산과 평가 행렬 처리 연산이 기하급수적으로 증가하는 문제이다. 이 문제는 처리 비용과도 연결되는데, 연산이 증가하면 속도는 느려지고 비용은 올라간다.

희박성 문제는 전체 상품 수에 비해 평가된 상품의 수가 극히 적을 경우 유사 사용자 탐색이 어려우며 추천의 질도 떨어진다는 문제이다.[3]

#### 2.2 협업 필터링

협업 필터링에 앞서 학사관리 시스템에 접근하여 데이터를 수집해야할 필요가 있다. 본 연구에서는 웹 크롤링 기술을 이용하여 학사관리 시스템의 데이터를 데이터베이스에 추가하려 한다. 먼저 크롬 개발자 도구를 이용하여 페이지의 요소의 이름(클래스 이름 등)을 추출한다. 다음으로 selenium 클래스의 find\_elements\_by\_class\_name() 함수에 자신이 찾고 싶은 클래스의 이름을 인자로 전달하고 함수를 호출한다. 만약 학교 홈페이지 학점 페이지의 성적 테이블로 접근하는 과정에서 프레임이 이중 프레임으로 되어 있다면, switch\_to.frame() 함수를 이용하여 원하는 테

이블로 이동한다. 이 방법은 웹사이트 내에서 모든 다중 프레임에서 이용할 수 있는 방법이다. 이렇게 추출한 테이블의 텍스트를 프로그램에 출력 저장한다. 이렇게 저장한 텍스트 값은 여러 함수를 이용하여 수정 제외할 수 있다.

이와 같은 방법의 장점으로 코드가 단순하고 가독성이 좋다는 점이다. 또한 코드의 난이도가 낮아 구축과 수정이 편리하다. 단점으로는 코드가 단순한 만큼 수동적인 코드이다. 홈페이지의 텍스트 값을 모를 경우 해당 프로그램은 작동되지 않거나 구축이 불가능하다. 그리고 모든 요소에 하나씩 접근하는 코드가기 때문에 시간복잡도가 큰 편에 속한다. 하지만 본 연구에서는 함수를 호출하는 횟수가 많지 않기 때문에 충분히 적합하다고 판단한다.

```
className = driver.find_elements_by_class_name('클래스네임')
myGrade = className['리스트 내 인덱스']
titles = ["이수학년도", "이수학기", "과목코드"]
for ch in titles:
    myGrade = myGrade.replace(ch, "")
```

(그림 1) web crawling codes

첫번째 줄 코드의 경우, 클래스에 접근하는 함수를 사용한 것이다. 이 함수의 장점은 가독성이 좋다는 것이지만 일일이 접근해야 한다는 단점이 있다. 또한 개발자가 직접 페이지에 접속하여 해당 요소를 확인하고, 그 요소가 가리키는 것이 정확히 무엇인지 판단하는 과정을 거쳐야 한다.

두번째 줄 코드는 driver.find\_elements\_by\_class\_name()의 인자를 가진 모든 클래스들이 리스트로 저장되어 반환되는데, 이 값을 myGrade 에 저장하는 것이다. 이 역시 인덱스에 저장된 값이 무엇인지 개발자가 직접 확인해야 정확한 코드 구현이 가능하다는 점에서 시간과 비용이 많이 소요될 수 있다.

□ 학기별 세부 성적

이수학년도	이수학기	과목코드
-------	------	------

(그림 2) Example of myGrade's titles

세번째 줄부터 다섯번째 줄의 코드에서 titles 에 저장된 리스트는 위의 내용이다. 이 내용은 숭실대학교 유세인트의 성적세부보기에 나와있는 표이다. 다른 학교의 데이터를 수집하려는 경우 타이틀이 다를 수 있으며 홈페이지별로 다른 titles 열 이름을 사용할 필요가 있다.

2.3 협업 필터링

웹 크롤링을 통해서 저장한 데이터를 이용한 협업 필터링 시스템을 제안한다. 이 시스템의 목표는 데이터를 분석하여 학생들에게 알맞은 과목을 추천하는 것에 있다. 따라서 적절한 기준에 따라 데이터를 분류하고 정렬하도록 설계한다.

- 수강년도 - 과목명 - 교수명

다음 세 가지의 기준에 따라서 사용자가 수강할 수 있는 과목에 대해서 그룹화한다. 예를 들면 한 그

룹은 [ “2020 년”, “자료구조”, “홍길동” ] 의 형식이라고 할 수 있다.

- 수강 인원(A) - 드랍 인원(D)
- 성적 분포(G) - 강의 평가(E)

다음 네 가지의 기준에 따라서 이전에 쌓인 데이터에 대해서 분석하고 정렬한다. 모든 기준에는 이전에 수강한 해당 년도의 전체 학생 수(N)에 대한 비율 또는 평균이 적용된다.

수강 인원의 비율은 A / N 이다. 또한 각 수강년도에 해당하는 수강 인원의 비율의 평균을 구한다. 구한 평균을 이용하여 내림차순으로 정렬한다.

드랍 인원의 비율은 D / N 이다. 또한 각 수강년도에 해당하는 드랍 인원의 비율의 평균을 구한다. 구한 평균을 이용하여 오름차순으로 정렬한다.

성적 분포는 수강한 인원의 성적을 학점을 기준으로 평균(Avg(G))을 구한다. 최대 4.5 점을 적용하는 학교의 기준으로 A+ = 4.5, A0 = 4.3, A- = 4.0 과 같은 학점을 이용하여 수강한 인원이 받은 성적의 평균을 구한다. 여기서 구한 평균을 각 수강년도에 적용하여 평균을 구한 후 내림차순으로 정렬한다.

강의 평가의 경우 평가받은 점수의 평균(Avg(E))을 구하고 각 수강년도에 적용하여 평균을 구한다. 그 후 내림차순으로 정렬한다.

이와 같이 다양한 정렬 기준을 만들고 그룹화한 데이터 각각에 접근하여 추천을 받을 수 있도록 한다. 예를 들면 같은 “과목명” 인 데이터 중 다른 “교수명”

인 데이터를 비교하여 평가가 좋은 강의를 추천받을 수 있다.

이와 같은 시스템의 장점은 아직 수강하지 않은 과목에 대해서 다른 이용자가 이용하여 기록되어 있는 데이터를 활용하여 선택에 도움을 줄 수 있다는 점이다. 단점으로는 데이터에 의존하기 때문에 주관적인 평가 및 만족도에 대해서는 알 수 없다는 것이다. 또한 이 시스템의 한계점으로는 해당 과목을 처음 강의 하게된 교수가 있을 경우, 축적된 과거의 데이터가 없기 때문에 분석이 불가능하다는 점이다. 이 단점은 협업 필터링의 초기 평가 문제점과도 연결된다.

3. 결론

현재 대부분의 대학에서 어떠한 과목에 대한 평가, 정보 등을 얻을 수 있는 것은 여전히 사람과 사람을 통해서이다. 주관적인 평가와 극히 일부분의 사람들이 전달해준 정보를 가지고 계속해서 합리적이지 않은 선택을 하게 된다는 것이다. 학생이 직접 본인이 수강할 과목을 선택할 권리가 있는 만큼 그 과목과 교수에 대한 정보도 알아야 한다고 생각했다. 하지만 현재 나와있는 시스템에서는 그런 정보를 알려주는 시스템이 거의 없고, 따라서 직접 데이터를 분석하게 되었다.

본 연구에서는 실생활에 도움이 될 것 같지만 아직 개발되지 않은 시스템에 대한 설계를 진행하였고, 실제로 시스템이 도입되었을 때 어떠한 효과를 얻을 수 있고, 학생과 학교에 어떠한 영향을 미치는가에

대해서 추가로 분석하고 수정하는 과정이 필요할 것이다. 학생들이 이 시스템을 이용함으로써 자신의 지식과 경력에 좋은 보탬이 될 것이라 기대한다.

#### 참고문헌

- [1] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, "An Algorithmic Framework for Performing Collaborative Filtering", Proc. of 22nd Annual International ACM SIGIR Conference, SIGIR Workshop on Recommender Systems, Berkeley, CA, Aug. 1999. Research and Development in Information Retrieval, pp. 230-237, Aug. 1999.
- [2] J. Konstan, D. B. Miller, D. Maltz, J. Herlocker, L. Gordon, and J. Riedl, "GroupLens: Applying collaborative filtering to Usenet news", Communications of ACM, Vol. 40, No. 3, pp. 77-87, March 1997.
- [3] Seok-Jong Yu, Integrated Preference Similarity Algorithm for Improving Sparsity Problem in Collaborative Filtering, Journal of KIIT. Vol. 11, No. 7, pp. 159-164, July 31, 2013.