

취업 큐레이션 시스템을 위한 개인 맞춤형 교육 콘텐츠 추천 기법

Personalized University Educational Contents Recommendation Scheme for Job Curation Systems

임종태*, 오영호**, 최재용*, 편도웅**, 이소민*, 신보경*, 채대성**, 복경수***, 유재수*
충북대학교 정보통신공학부*, 충북대학교 빅데이터학과**, 원광대학교 SW융합학과***

Jongtae Lim(jtlim@chungbuk.ac.kr)*, Youngho Oh(ohy5268@kakao.com)**,
JaeYong Choi(headmeat@naver.com)*, DoWoong Pyun(pyun19@naver.com)**,
Somin Lee(somin@chungbuk.ac.kr)*, Bokyoung Shin(sbk02@chungbuk.ac.kr)*,
Daesung Chae(2016178002@chungbuk.ac.kr)**, Kyoungsoo Bok(ksbok@wku.ac.kr)***,
Jaesoo Yoo(yjs@chungbuk.ac.kr)*

요약

최근 모바일 기기 및 소셜 미디어 서비스의 발전으로 인해, 콘텐츠 추천 기법에 대한 연구들이 진행되고 있다. 그러한 콘텐츠 추천 기법들은 일반적으로 취업 큐레이션 시스템에 적용된다. 대부분의 기존 대학 교육 콘텐츠 추천 기법은 해당 학생의 학교, 전공 등을 바탕으로 가장 많이 수강한 과목만을 추천하기 때문에 각 학생이 원하는 취업의 형태나 분야 등이 고려되지 않는다. 본 논문에서는 취업 큐레이션 시스템 기반의 개인 맞춤형 교육 콘텐츠 추천 기법을 제안한다. 제안하는 기법은 사용자가 취업 큐레이션 시스템에서 활동한 내용을 바탕으로 사용자가 원하는 취업 형태와 취업 분야를 분석하여 관심 기업을 도출한다. 제안하는 기법은 도출된 관심 기업을 바탕으로 관심 기업에 취업한 졸업생들의 신뢰도와 유사도를 측정하여 참고할만한 졸업생을 선정하고 협업 필터링을 통해 사용자에게 맞춤형 교과목, 비교과목, 자율 활동 목록을 추천한다.

■ 중심어 : | 취업 큐레이션 | 교육 콘텐츠 | 콘텐츠 추천 | 맞춤형 추천 | 추천 시스템 |

Abstract

Recently, with the development of mobile devices and social media services, contents recommendation schemes have been studied. They are typically applied to the job curation systems. Most existing university education content recommendation schemes only recommend the most frequently taken subjects based on the student's school and major. Therefore, they do not consider the type or field of employment that each student wants. In this paper, we propose a university educational contents recommendation scheme for job curation services. The proposed scheme extracts companies that a user is interested in by analyzing his/her activities in the job curation system. The proposed scheme selects graduates or mentors based on the reliability and similarity of graduates who have been employed at the companies of interest. The proposed scheme recommends customized subjects, comparative subjects, and autonomous activity lists to users through collaborative filtering.

■ keyword : | Job Curation | Educational Contents | Contents Recommendation | Personalized Recommendation | Recommendation System |

* 이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(No. 2019R1A2C2084257), 중소벤처기업부 '산업전문 인력역량강화사업'의 재원으로 한국산학협력회(AURI)의 지원(2021년 기업연계형연구개발인력양성사업, 과제번호 : S3047889), 과학기술정보통신부 및 정보통신기획평가원의 지역지능화혁신인재양성(Grand ICT연구센터) 사업(IITP-2021-2020-0-01462) 그리고 2020년 대한민국 교육부와 한국연구재단의 지원(NRF-2020S1A5B8103910)을 받아 수행된 연구임.

* 본 논문은 ICC2020 국제학술대회 우수논문입니다.

접수일자 : 2021년 05월 20일

수정일자 : 2021년 06월 11일

심사완료일 : 2021년 06월 11일

교신저자 : 유재수, e-mail : yjs@chungbuk.ac.kr

I. 서론

최근 스마트폰, 태블릿PC, Laptop 등의 모바일 기기 및 소셜 미디어 서비스의 발전으로 인해 데이터의 양이 급격히 증가하고 있다. 이로 인해 엄청난 규모의 정보 중에서 사용자에게 의미가 있을만한 정보를 추천하는 다양한 추천 기법이 중요하게 연구되고 있다[1-8]. 추천 기법은 우리가 사용하는 다양한 서비스에 활용되고 있다. 콘텐츠 추천 기법은 취업 큐레이션 시스템에도 적용되고 있다. 취업 큐레이션 시스템은 대학 교육 과정에 있는 학생들을 대상으로 취업에 도움이 될 만한 교과목, 비교과목, 어학, 현장실습, 봉사활동 등과 같은 다양한 활동을 추천한다. 하지만 대부분의 기존 취업 큐레이션 시스템은 해당 학생의 학교, 전공 등을 바탕으로 가장 많이 수강한 과목을 추천하기 때문에, 각 학생이 원하는 취업의 형태나 원하는 분야는 고려되지 않는다. 각 학생이 수강하거나 수행할 수 있는 활동은 매우 다양하게 존재하기 때문에, 각 학생은 자신에게 적합한 교과목, 비교과목, 자율 활동을 선택하는데 어려움을 겪는다. 따라서 취업 큐레이션 시스템은 각 학생이 졸업 후에 자신이 원하는 직장에 취업하기 위하여 어떤 기술이나 경험이 필요한지를 고려하여 교과목, 비교과목, 자율 활동을 추천하는 대학 교육 콘텐츠 추천 기법이 필요하다. 전통적으로 콘텐츠 추천 기법은 두가지 접근 방법이 존재한다. 첫 번째는 아이템 기반의 추천이다[1]. 아이템 기반의 추천 기법은 사용자가 과거에 경험했던 아이템을 기반으로 추천을 수행한다. 이 기법은 사용자가 많이 경험한 아이템이나 경험을 한 뒤에 높은 평점을 부여한 아이템과 유사한 속성을 가진 아이템을 추천하는 콘텐츠 추천 기법이다. 하지만 아이템 기반의 콘텐츠 추천 기법은 항상 비슷한 아이템만을 추천하기 때문에 사용자가 경험하지 못한 속성을 가진 아이템들에 대한 추천이 불가능하여 다양한 경험을 하도록 유도하지 못한다는 단점이 존재한다. 두 번째는 협업 필터링(Collaborative Filtering)이다[2-8]. 협업 필터링은 사용자의 경험과 유사한 다른 사용자를 검색하고 유사한 다른 사용자들이 많이 경험했거나 좋은 평점을 부여했지만 사용자가 경험하지 못한 아이템을 추천한다[5][6]. 협업 필터링을 이용한 추천의 정확도를

향상시키기 위해서는 유사한 사용자를 잘 검색하는 방법이 중요한데, 이는 협업 필터링이 사용되는 응용에 맞게 설계되어야 한다[2][3][7].

본 논문에서는 취업 큐레이션 시스템 기반의 개인 맞춤형 교육 콘텐츠 추천 기법을 제안한다. 제안하는 기법은 사용자가 취업 큐레이션 시스템에서 활동한 내용을 바탕으로 사용자가 원하는 취업 형태와 취업 분야를 분석하여 관심 기업을 도출한다. 제안하는 기법은 이렇게 도출된 관심 기업을 바탕으로 관심 기업에 취업한 졸업생들의 신뢰도와 유사도를 측정하여 참고할만한 졸업생을 선정하고 협업 필터링을 통해 사용자에게 맞춤형 교과목, 비교과목, 자율 활동 목록을 추천한다.

본 논문의 구성은 다음과 같다. 2장에서는 기존 교육 콘텐츠 추천 기법을 분석하고 기존 기법들의 한계를 기술한다. 3장에서는 제안하는 대학 교육 콘텐츠 추천 기법을 기술한다. 4장에서는 제안하는 대학 교육 콘텐츠 추천 기법을 기반으로한 교육 콘텐츠 추천 시스템의 구현 내용을 기술한다. 5장에서는 제안하는 콘텐츠 추천 기법의 우수성을 확인하기 위하여 성능평가를 수행한 결과를 기술한다.

II. 관련 연구

교육 콘텐츠 추천을 위한 다양한 기법들이 연구됐다[9-15]. 이용준[9] 등은 전자상거래에서 이용되던 협업 필터링을 교육용 콘텐츠 추천에 활용하는 연구를 수행했다. 이 기법은 각 참여자가 교육 콘텐츠에 평가점수를 부여하고 평가점수를 기반으로 유사한 사용자를 검색하여 협업 필터링을 수행하는 교육 콘텐츠 추천기법과 유사도 보정 방법 등을 제안했다. 이운수[10] 등은 교육 콘텐츠를 검색하기 위하여 역량 온톨로지라는 개념을 사용했다. 이 기법은 학습자에게 맞춤형 콘텐츠를 추천하는데 가장 중요한 요소를 학습자의 역량이라고 생각하여, 학습자의 역량 정보와 목표를 나타내는 포트폴리오와 과거 학습자료(리스트) 데이터를 활용하여 교육 콘텐츠를 추천한다. R. Noor[11] 등은 사용자에게 맞춤형 추천을 수행하기 위한 하이브리드(Hybrid) 추천 기법을 제안했다. 이 기법은 사용자가 경험한 유사

한 콘텐츠, 사용자의 경험과 유사한 사용자, 사용자의 학습 스타일을 종합적으로 고려한 교육 콘텐츠 추천 기법을 설계했다. 유사한 콘텐츠를 분석함으로써 사용자가 흥미있어하는 주제의 교육 콘텐츠를 도출하며, 유사한 사용자로부터 협업 필터링을 통한 분석을 수행함으로써 새로운 교육 콘텐츠를 도출하여 추천할 수 있다. 이 기법은 두 가지 고려사항에 사용자의 학습 스타일을 반영하여 최종 교육 콘텐츠를 추천한다. L. G. Ferreira[15] 등은 학습자 그룹을 기반으로한 교육 콘텐츠 추천 기법을 제안했다.

이후에도 몇몇의 교육 콘텐츠 추천을 위한 다양한 기법들이 연구됐다[9-15]. 하지만 기존 기법들은 대부분 교육 콘텐츠에 사용자가 부여하는 평점이 존재하고 이를 이용하여 유사한 콘텐츠를 추천하거나 협업 필터링을 수행하여 교육 콘텐츠를 추천한다. 하지만 본 논문에서 기반을 두고 있는 취업 큐레이션 시스템으로부터 생성되는 데이터 중에는 교과목과 같이 점수가 존재하는 콘텐츠 외에도 비교과목, 자율활동과 같이 명시적 평가가 존재하지 않는 콘텐츠도 존재한다. 이러한 교육 콘텐츠에 대한 추천 결과를 만들어내기 위해서는 사용자가 직접 부여하는 평가 점수 외에도 좋아요, 즐겨찾기, 댓글의 긍/부정 등 사용자의 소셜 행위를 분석하여 해당 콘텐츠의 선호 여부를 계산하는 것이 필요하다. 또한 기존 기법들이 양질의 교육 콘텐츠를 추천하는 것 그 자체가 목적인 것과 달리, 본 논문에서 제안하는 콘텐츠 추천 기법은 취업 큐레이션 시스템에 활용될 추천 기법이기 때문에, 사용자가 선호하는 취업 형태나 기업 등을 고려하여 교육 콘텐츠를 추천하는 것이 매우 중요하다.

III. 제안하는 개인 맞춤형 교육 콘텐츠 추천 기법

1. 제안하는 기법의 특징

본 논문에서는 취업 큐레이션 시스템 기반의 맞춤형 대학 교육 콘텐츠 추천 기법을 제안한다. 본 논문에서 기반이 되는 취업 큐레이션 시스템에서는 대학 교육 과정에 있는 학생들이 수강한 교과목, 비교과목, 그리고

어학, 현장실습, 봉사활동 등과 같은 학생들의 활동 내역이 저장된다. 또한 기업은 취업 정보 등을 시스템에 업로드 하는데 이를 학생들이 조회하거나 지원한 정보가 저장된다. 또한 기타 취업과 관련된 정보를 페이스북이나 유튜브 등의 소셜 미디어 서비스에서 검색 및 조회한 정보가 저장된다. 제안하는 추천 기법은 이러한 취업 큐레이션 시스템을 기반으로 저장된 학생 활동 및 취업 정보들을 분석하여 개인 맞춤형 교육 콘텐츠를 추천한다. [그림 1]은 제안하는 추천 시스템의 전체 구조를 보여준다. 제안하는 콘텐츠 추천 기법은 데이터 수집 및 저장 관리를 기반으로 관심 기업 예측, 신뢰도 / 유사도 분석, 추천 결과 생성 과정으로 구성된다.

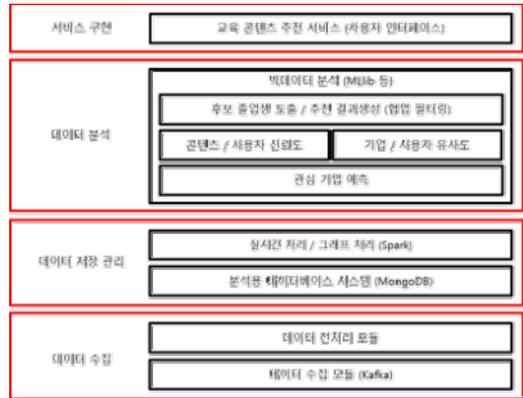


그림 1. 제안하는 대학 교육 콘텐츠 추천 시스템 구조

2. 관심 기업 예측

관심 기업 예측은 각 학생의 저장된 취업 관련 정보를 바탕으로 각 사용자가 원하는 기업의 형태나 분야를 예측한다. [그림 2]는 관심 기업 예측 모듈을 보여준다. 채용 신청 정보는 기업이 등록한 취업 정보에 사용자가 취업을 위하여 지원을 한 정보이다. 채용 신청 정보는 사용자가 직접 신청한 정보이기 때문에 사용자 분석에 매우 높은 비중을 차지한다. 채용 검색 정보는 대학 교육 콘텐츠 큐레이션 시스템 및 소셜 미디어 서비스에서 검색을 수행한 이력과 검색 결과 중에 조회한 콘텐츠 / 기업에 대한 이력 정보이다. 하지만 해당 콘텐츠를 열람 후 만족도를 입력받지 않으면 사용자가 만족한 콘텐츠인지 알 수 없기 때문에 사용자 분석에 약간의 비중을 차지한다. 해당 정보들을 분석하여 제안하는 추천

기법은 사용자가 원하는 기업의 형태나 분야, 지역 등을 기반으로 가상의 관심 기업을 도출한다. 만약 채용 신청 정보나 채용 검색 정보가 분석을 수행할 만큼 충분하지 않다면 사용자의 학교, 학과, 전공, 기타 개인정보들을 분석하여 대중성 있는 일반적인 기업을 관심 기업으로 예측한다.



그림 2. 관심기업 예측 모듈

3. 유사도 분석

제안하는 기법에서는 두 가지 유사도 분석이 요구된다. 첫 번째 유사도 분석으로 제안하는 기법은 각 사용자의 예측된 가상의 관심 기업을 바탕으로 실제 존재하는 기업 중에 유사한 기업을 분석한다. 두 번째 유사도 분석으로 제안하는 기법은 각 사용자가 수강한 교과목, 비교과목, 자율 활동을 기반으로 유사한 대학 교육 과정을 경험한 졸업생을 분석한다. 기업 유사도 분석 결과 및 사용자 유사도 분석 결과는 참고할만한 졸업생을 도출하는데 활용된다. 제안하는 기법은 유사도 분석을 위해 기업 유사도 분석에서는 기업의 형태, 분야, 지역 등이 벡터 값으로 사용되며, 사용자 유사도 분석에서는 사용자의 교과목, 비교과목, 자율활동 각각에 수강여부 및 만족도 등이 벡터 값으로 사용된다. 이러한 값들은 기업 매트릭스 및 사용자 활동 매트릭스로 구축되어 데이터베이스에 저장된다. [그림 3]은 활동 / 점수 매트릭스를 작성하는 과정을 보여준다. 제안하는 기법에서는 작성된 매트릭스를 기반으로 피어슨 상관계수[16]를 활용한 유사도 분석을 수행하여 기업 간 / 사용자 간의 유사도를 계산한다.

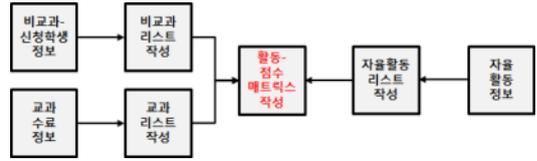


그림 3. 활동-점수 매트릭스 작성

4. 신뢰도 분석

제안하는 기법에서는 두 가지 신뢰도 분석이 요구된다. 첫 번째 신뢰도 분석으로 제안하는 기법은 각 사용자가 얼마나 취업 큐레이션 시스템을 적극적으로 활용하였는지를 분석한다. 사용자 신뢰도 분석 결과는 참고할만한 졸업생을 도출하는데 활용된다. 두 번째 신뢰도 분석으로 제안하는 기법은 각 콘텐츠에 부여된 평점과 같은 명시적 평가 점수와 댓글의 긍/부정, 좋아요, 즐겨찾기, 싫어요, 동영상의 재생시간, 해당 포스트에 머문 시간 등의 암시적 평가를 종합하여 해당 콘텐츠가 유일한 콘텐츠인지를 판단한다. [그림 4]는 콘텐츠 신뢰도 분석 모듈을 보여준다. 콘텐츠 신뢰도는 추후 최종 추천 결과를 생성하는데 활용된다. 취업 큐레이션 시스템에서는 교과목, 비교과목에 대하여 수강한 학생들이 만족도와 댓글을 남길 수 있기 때문에 해당 내용을 기반으로 콘텐츠의 신뢰도를 평가하며, 소셜 미디어 서비스를 통한 대학 교육 콘텐츠에서는 앞서 언급한 다양한 소셜 행위를 기반으로 신뢰도를 판단한다.



그림 4. 콘텐츠 신뢰도 분석 모듈

5. 후보 졸업생 도출

제안하는 기법에서는 대학 교육 콘텐츠 추천을 위하여 참고할만한 졸업생을 도출한다. 참고할만한 졸업생은 신뢰도와 유사도를 결합하여 결정한다. [그림 5]는 후보 졸업생을 도출하는 과정을 보여준다. 첫 번째, 기업 유사도는 해당 졸업생이 취업한 기업 및 사용자의 관심 기업의 유사한 정도이다. 기업 유사도는 수식 1로

계산한다. 수식에서 $Sim_{comp}(i, j)$ 은 사용자 i 와 졸업생 j 의 기업 유사도 결과이다. 이는 사용자의 관심 기업과 유사한 기업에 취업한 졸업생을 참고하는 것이 대학 교육 콘텐츠 추천에 유의미하다는 것을 의미한다. 두 번째, 사용자 유사도는 해당 졸업생이 사용자와 얼마나 비슷한 경험을 수행하였는지를 의미한다. 사용자 유사도는 수식 2로 계산한다. 수식에서 $Sim_{user}(i, j)$ 는 사용자 i 와 졸업생 j 의 사용자 유사도 결과이다. 이는 비슷한 경험을 수행한 졸업생을 참고하는 것이 추후 수강할 교과목, 비교과목, 자율 활동 등을 선택하여 실제 수강하는데 흐름을 크게 바꾸지 않아도 되어 유리하다는 것을 의미한다. 세 번째, 사용자 신뢰도는 본 논문에서 기반이 되는 취업 큐레이션 시스템을 적극적으로 활용한 정도이다. 신뢰도는 수식 3으로 계산한다. 이는 취업 큐레이션 시스템에서 활동한 정보가 많은 졸업생일수록 보다 정확한 분석 결과를 도출할 수 있기 때문에 대학 콘텐츠 추천에 유리하다는 것을 의미한다.

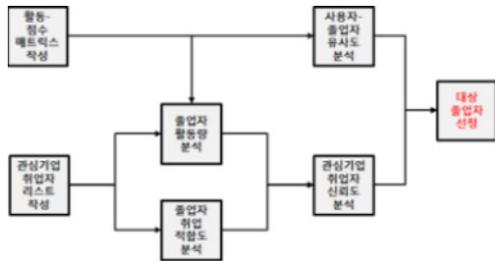


그림 5. 후보 졸업생 도출

제안하는 기법은 3가지 고려 사항을 결합하여 도출된 값으로 Top-k 질의를 수행한 결과인 k명의 졸업생을 반환한다. Top-k 질의는 수식 4를 통해 수행한다. 수식에서 α, β, γ 는 보정 상수로 각 유사도 및 신뢰도를 같은 범위의 값으로 만들어 동등하게 비교될 수 있도록 한다. 그리고 w_1, w_2, w_3 는 각 고려사항을 얼마만큼의 비중으로 중요하게 생각할 것인지를 나타내는 가중치로 사용자나 서비스에 따라 설정할 수 있다. 제안하는 기법은 수식 4를 통해 계산되는 값을 기준으로 가장 높은 값을 가진 졸업생들을 선택하여 목록을 생성한다.

$$Sim_1 = \alpha * w_1 * Sim_{comp}(i, j) \quad (1)$$

$$\alpha = 1 / MAX(Sim_{comp})$$

$$Sim_2 = \beta * w_2 * Sim_{user}(i, j) \quad (2)$$

$$\beta = 1 / MAX(Sim_{user})$$

$$Trust_1 = \gamma * w_3 * Trust_{user}(i, j) \quad (3)$$

$$\gamma = 1 / MAX(Trust_{user})$$

$$Value(i, j) = Sim_1(i, j) + Sim_2(i, j) + Trust_1(i, j) \quad (4)$$

$$w_1 + w_2 + w_3 = 1$$

6. 최종 결과 생성

제안하는 기법에서는 참고할만한 졸업생 목록을 기준으로 협업 필터링을 수행하여 최종 콘텐츠를 추천한다. [그림 6]은 추천 결과를 생성하는 과정을 보여준다. 협업 필터링 결과, 사용자는 수강하지 않았지만 참고할만한 졸업생들은 많이 수강하였고, 만족도가 높았던 교과목, 비교과목, 자율 활동의 목록이 생성된다. 하지만 이렇게 생성한 목록은 졸업생이 부여한 명시적 평가인 만족도만을 기준으로 작성되기 때문에, 콘텐츠 신뢰도를 판단하기 위한 정보가 부족한 경우가 많다. 따라서 본 논문에서는 신뢰도 분석에서 언급한 암시적 평가를 종합하여 콘텐츠의 신뢰도를 판단하여 최종 결과를 보정한다.



그림 6. 추천 결과 생성 모듈

콘텐츠 신뢰도는 수식 5로 계산된다. 제안하는 기법에서는 콘텐츠 신뢰도를 가중치로 사용하여, 적용된 최종 점수가 각 콘텐츠별로 계산된다. 최종 점수는 수식 6으로 계산된다. 수식에서 δ, ϵ 는 보정 상수이며, w_4 는 가중치이다. 최종 계산된 콘텐츠별 점수를 바탕으로 제안하는 기법은 사용자에게 추천할 교과목, 비교과목, 자율 활동 목록을 생성한다.

$$Trust_2 = w_4 * Trust_{cont} \quad (5)$$

$$\delta = 1 / MAX(Trust_{cont})$$

$$\text{Result}_{\text{cont}} = (1 - w_4) * \epsilon * \text{Rating}_{\text{cont}} * \text{Trust}_2 \quad (6)$$

$$\epsilon = 1 / \text{MAX}(\text{Rating}_{\text{cont}})$$

IV. 제안하는 콘텐츠 추천 시스템 구현

[그림 7]은 제안하는 대학 교육 콘텐츠 추천 시스템의 전체 구조를 보여준다. 제안하는 시스템은 오라클 DB에 구축된 취업 큐레이션 시스템 사용 데이터를 기반으로 한다. 제안하는 시스템은 취업 큐레이션 시스템 사용 데이터 중에서 분석에 필요한 정보를 실시간으로 수집하기 위하여 아파치 카프카(Apache Kafka)를 이용한다. 수집된 데이터는 전처리 후에 몽고 DB(Mongo DB)에 저장되어 유사도 분석 및 신뢰도 분석에 활용된다. 제안하는 시스템은 유사도 분석 및 신뢰도 분석을 수행하기 위하여 분산 병렬 처리 플랫폼인 아파치 스파크(Apache Spark)를 사용한다. 제안하는 추천 시스템은 설계된 추천 기법에 따라 결과를 생성한다. 생성된 결과는 GUI 기반의 사용자 인터페이스를 통해 서비스 된다.



그림 7. 제안하는 대학 교육 콘텐츠 추천 시스템

데이터 수집기는 대학 교육 콘텐츠 추천을 위하여 사용자들이 직접 수강하거나 활동한 교과목, 비교과목, 자율 활동 정보 등 오프라인 활동 정보를 수집한다. 또한 데이터 수집기는 각 사용자들이 대학 교육 콘텐츠 큐레이션 시스템에서 교과목 / 비교과목의 평점, 댓글, 취업 지원, 조회, 콘텐츠 검색 등 온라인 활동 정보를 수집한다. [그림 8]은 데이터 수집기를 통해 수집한 사용자 활동 정보 테이블을 보여준다. 데이터 수집기는 아파치 카프카 플랫폼을 활용했고 파이썬 프로그래밍 언어로 구현했다. 이렇게 수집된 온라인과 오프라인 활동 정보

는 데이터 저장 및 분석 서버의 데이터베이스 관리 시스템인 몽고 DB에 저장 관리된다.

```
[root@localhost consumer]# ls
CPS_BOARD_REPLY.py      CPS_SCHOOL_EMPLOY_INFO.py
CPS_CODE_MNG.py         CPS_SCHOOL_EMPLOY_STD_INFO.py
CPS_EMPLOY_SEARCH_HIS.py CPS_STAR_POINT.py
CPS_GRADUATE_CORP_INFO.py exec.py
CPS_NCR_PROGRAM_INFO.py func.py
CPS_NCR_PROGRAM_STD.py  py
CPS_OUT_ACTIVITY_MNG.py pycache
```

그림 8. 수집된 사용자 활동 정보 테이블

콘텐츠 추천 엔진은 3장에서 설계된 콘텐츠 추천 기법을 기반으로 동작한다. 콘텐츠 추천 엔진은 실제 유사도 / 신뢰도 분석을 수행하고 추천 결과를 만들어 사용자 인터페이스에 표시하기 위한 데이터를 생성한다. [그림 9]는 추천 기법을 수행한 결과로 저장된 교과목 예시를 보여준다. 콘텐츠 추천 엔진은 아파치 스파크 플랫폼을 활용했고 스텝 프로그래밍 언어로 구현했다. 콘텐츠 추천 엔진은 추천 결과를 생성하여 데이터 저장 및 분석 서버의 데이터베이스 관리 시스템인 몽고 DB에 문서형 형태로 저장한다. 제안하는 추천 시스템은 사용자가 취업 큐레이션 시스템에 로그인하면 추천에 필요한 유사도 / 신뢰도 분석을 미리 수행하여 추천 결과를 만들어 DB에 저장한다. 제안하는 콘텐츠 추천 시스템은 사용자가 취업 큐레이션 시스템을 사용하다가 사용자 인터페이스를 통해 분석 결과를 요청할 경우, 미리 저장된 추천 결과를 보여줌으로써 사용자가 빠르게 추천 결과를 열람할 수 있도록 한다.

```
{
  "_id" : ObjectId("5ea2773b43b81e5009951a6f"),
  "Rank" : 1,
  "SBJT_KEY_CD" : "CAC00621"
}
{
  "_id" : ObjectId("5ea2773b43b81e5009951a73"),
  "Rank" : 2,
  "SBJT_KEY_CD" : "AAM00331"
}
```

그림 9. 콘텐츠 추천 결과 예시

V. 성능평가

본 논문에서는 제안하는 추천 기법의 우수성을 확인하기 위하여 추천 결과의 정확도를 측정했다. [표 1]은

성능평가 환경을 보여준다. 정확도는 각각 정밀도(Precision), 재현율(Recall), F-measure로 측정했다. 성능평가에서 추천의 정답셋은 사용자에게 대하여 일정 기간 이전의 활동을 분석하여 추후에 추천된 교과목, 비교과목, 기타 활동을 남은 기간 동안 실제로 수행하였는지를 이용하여 간접적으로 정확도를 측정했다. 제안하는 콘텐츠 추천 기법은 기반이 되는 취업 큐레이션 시스템의 사용 데이터에 맞게 만들어진 기법이기 때문에 기존 콘텐츠 추천 기법과의 비교가 불가능하여 자체 정확도에 대한 평가만을 수행했다.

표 1. 성능평가 환경

항목	값
환경	CentOS7 / Apache Spark / Scala
DB	Mongo DB
데이터	학교 20개, 학생 500명, 기업 300개 (샘플)
정답셋	30개
추천 개수	10-60개
평가항목	정확도 (정밀도, 재현율, F-measure)

[그림 10]은 추천 결과 개수 변화에 따른 정밀도 측정 결과를 보여준다. 성능평가 결과, 제안하는 추천 기법은 70% 이상의 정밀도를 보임을 확인했다. [그림 10]에서 정밀도는 추천 개수가 40개일때부터 점차 낮아지기 시작하여 이후 급격히 낮아지는 경향을 보인다. 그 이유는 정답셋에 해당되는 30개의 결과가 최상위 30개 안에 대부분 존재하고 50개 이후에는 존재하지 않는데, 성능평가에는 추천 개수를 지속적으로 증가시켜가며 측정했기 때문으로 해석된다.

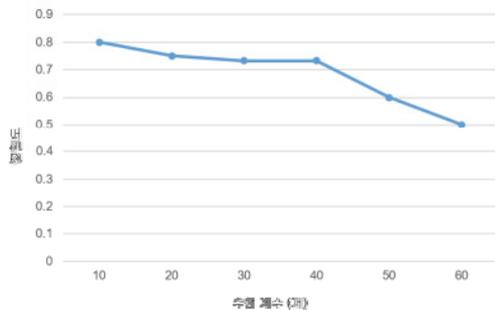


그림 10. 추천 개수에 따른 정밀도 측정 결과

[그림 11]은 추천 결과 개수 변화에 따른 재현율 측정 결과를 보여준다. 성능평가 결과, 제안하는 추천 기법은 70% 이상의 재현율을 보임을 확인했다. [그림 11]에서 처음 재현율은 정답셋에 비해 추천 개수가 현저히 적기 때문에 27%의 낮은 수치를 보이다가 추천 개수가 증가함에 따라 점차 증가한다. 성능평가 결과, 제안하는 추천 기법은 추천 개수가 40일 때 거의 모든 정답셋을 포함하며, 추천 개수가 50일 때 모든 정답셋을 결과에 포함하고 있음을 확인할 수 있다.

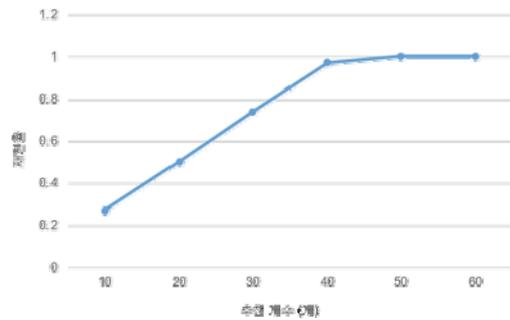


그림 11. 추천 개수에 따른 재현율 측정 결과

[그림 12]는 제안하는 콘텐츠 추천 기법의 종합적인 정확도를 보여준다. 추천 기법이 얼마나 정확한지를 나타내는 정밀도와 재현율은 파라미터 세팅에 따라 trade-off 관계를 보이므로 두 가지 정확도가 병합된 결과를 F-Measure로 도출했다. F-Measure는 수식 7로 계산한다. 성능평가 결과, 평균 66.53%의 정확도를 보였다. 본 성능평가 결과를 통해 정답셋이 30개일 경우, 추천 개수 30~40개 구간에서 제안하는 콘텐츠 추천 기법의 성능이 제일 우수한 것을 확인했으며, 이를 고려하여 취업 큐레이션 시스템에 활용할 예정이다.

$$F\text{-measure} = 2 * (\text{Precision} * \text{Recall} / (\text{Precision} + \text{Recall})) \quad (7)$$

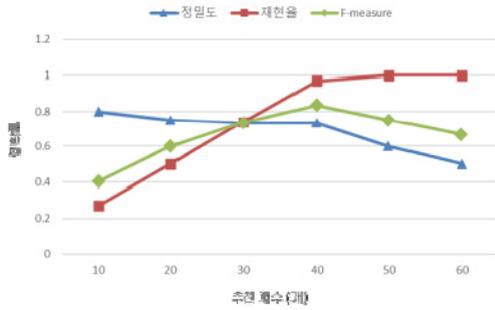


그림 12. 추천 개수 변화에 따른 정확도

VI. 결론

본 논문에서는 취업 큐레이션 시스템 기반의 개인 맞춤형 교육 콘텐츠 추천 기법을 제안했다. 제안하는 기법은 사용자가 취업 큐레이션 시스템에서 활동한 내용을 바탕으로 사용자가 원하는 취업 형태와 취업 분야를 분석하여 관심 기업을 도출한다. 제안하는 기법은 이렇게 도출된 관심 기업을 바탕으로 관심 기업에 취업한 졸업생들의 신뢰도와 유사도를 측정하여 참고할만한 졸업생을 선정하고 협업 필터링을 통해 사용자에게 맞춤형 교과목, 비교과목, 자율 활동 목록을 추천한다. 성능평가 결과, 평균 66.53%의 정확도를 보였다. 또한 30~40개의 추천 개수에서 가장 정확도가 높은 결과를 보였으므로 이를 고려하여 취업 큐레이션 시스템에 활용 예정이다.

추후 연구로는 더 다양한 정확도 및 처리 시간 측면의 성능 평가를 수행하여 제안하는 기법이 우수함을 보일 예정이다. 또한 현재 추천을 위한 유사도 / 신뢰도 분석에서 다수 대 다수 비교가 빈번히 발생하기 때문에 학생, 기업, 콘텐츠가 증가할 경우, 비교 연산이 기하급수적으로 증가하여 매우 큰 복잡도를 가진다. 따라서 추후에는 유사도 / 신뢰도 분석에서 비교 연산을 효율적으로 수행하여 복잡도를 감소시키기 위한 콘텐츠 추천 기법을 연구할 예정이다. 마지막으로 기존 큐레이션 시스템과의 만족도 평가를 통해 콘텐츠 추천 시스템의 우수성을 보일 예정이다.

참고 문헌

- [1] B. M. Sarwar, G. Karypis, J. A. Konstan and J. Riedl, "Item-based collaborative filtering recommendation algorithms," *Www*, Vol.1, pp.285-295, 2001.
- [2] Z. Zhao and M. Shang, "User-based collaborative-filtering recommendation algorithms on hadoop," *Proc. of the International Conference on Knowledge Discovery and Data Mining*, 2010.
- [3] L. Yao, Z. Xu, X. Zhou, and B. Lev, "Synergies Between Association Rules and Collaborative Filtering in Recommender System: An Application to Auto Industry," *Data Science and Digital Business*, pp.65-80, 2019.
- [4] J. L. Herlocker, J. A. Konstan, and J. Riedl, "Explaining collaborative filtering recommendations," *Proc. of the 2000 ACM conference on Computer supported cooperative work*, pp.241-250, 2000.
- [5] X. He, L. Liao, H. Zhang, L. Nie, X. Hu, and T. S. Chua, "Neural collaborative filtering," *Proc. of the 26th international conference on world wide web*, pp.173-182, 2017.
- [6] X. Su and T. M. Khoshgoftaar, "A survey of collaborative filtering techniques," *Advances in artificial intelligence*, 2009.
- [7] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl, "GroupLens: Applying collaborative filtering to usenet news," *Communications of the ACM*, Vol.40, No.3, pp.77-87, 1997.
- [8] Y. Koren and R. Bell, "Advances in collaborative filtering," *Recommender systems handbook*, pp.77-118, 2015.
- [9] Y. Lee, S. Lee, and C. Wang, "The Educational Contents Recommendation System Design based on Collaborative Filtering Method," *Journal of Korean Association of Computer Education*, Vol.6, No.2, pp.147-156, 2003.
- [10] Y. S. Lee, B. C. Chang, H. S. Kang, and J. H. Cha, "The educational contents recommendation

system using the competency ontology,” Journal of Digital Contents Society, Vol.11, No.4, pp.487-494, 2010.

- [11] R. Noor and A. K. Farman, “Personalized recommendation strategies in mobile educational systems,” Proc. of International Conference on Innovative Computing Technology (INTECH). IEEE, 2016.
- [12] N. Saraswat, H. Ghosh, M. Agrawal, and U. Narayanan, “Contextual recommendation of educational contents,” Proc. of International Conference on Artificial Intelligence in Education, pp.439-448, 2015.
- [13] P. Di Bitonto, T. Roselli, and V. Rossano, “Recommendation in e-learning social networks,” Proc. of International Conference on Web-Based Learning, pp.327-332, 2011.
- [14] C. Romero, S. Ventura, J. A. Delgado, and P. De Bra, “Personalized links recommendation based on data mining in adaptive educational hypermedia systems,” Proc. of European conference on technology enhanced learning, pp.292-306, 2007.
- [15] L. G. Ferreira, J. L. V. Barbosa, J. C. Gluz, V. K. Matter, and D. N. F. Barbosa, “Using Learner Group Profiles for Content Recommendation in Ubiquitous Environments,” International Journal of Information and Communication Technology Education (IJICTE), Vol.16, No.4, pp.1-19, 2020.
- [16] H. V. Nguyen and L. Bai, “Cosine similarity metric learning for face verification,” In Asian conference on computer vision, pp.709-720, 2010.

저 자 소 개

임 종 태(Jongtae Lim)

정회원



- 2009년 2월 : 충북대학교 정보통신 공학과(공학사)
 - 2011년 2월 : 충북대학교 정보통신 공학과(공학석사)
 - 2015년 8월 : 충북대학교 정보통신 공학과(공학박사)
 - 2015년 9월 ~ 2019년 8월 : 충북대학교 정보통신공학과 Postdoc.
 - 2019년 10월 ~ 현재 : 충북대학교 전자정보대학 정보통신공학부 초빙부교수
- 〈관심분야〉 : 소셜 미디어, 빅데이터, 인텔리전트 데이터베이스, 시공간 데이터베이스, 위치기반 서비스 등

오 영 호(Youngho Oh)

준회원



- 2015년 8월 : 충북대학교 공업화학 과(공학사)
- 2020년 3월 ~ 현재 : 충북대학교 빅데이터 협동과정 석사과정

〈관심분야〉 : 연속 질의 처리, 그래프 스트림, 빅데이터

최 재 용(Jaeyong Choi)

준회원



- 2020년 2월 : 충북대학교 정보통신 공학부(공학사)
- 2020년 3월 ~ 현재 : 충북대학교 정보통신공학부 석사과정

〈관심분야〉 : 소셜 네트워크, 빅데이터 처리 등

편 도 응(DoWoong Pyun)

준회원



- 2020년 2월 : 가천대학교 산업경영 공학과 (공학사)
- 2020년 3월 ~ 현재 : 충북대학교 빅데이터 협동과정 석사과정

〈관심분야〉 : 사물 인터넷, 빅데이터, 데이터베이스, 위치 기반 서비스 등

이 소 민(Somin Lee)

준회원



- 2019년 2월 : 충북대학교 정보통신 공학부 (공학사)
- 2019년 9월 ~ 현재 : 충북대학교 정보통신공학과 석사과정

〈관심분야〉 : 그래프 처리, 연속 질의 처리, 빅데이터, 기계 학습

신 보 경(Bokyoung Shin)

준회원



- 2016년 2월 : 충북대학교 정보통신 공학부(공학사)
- 2019년 3월 ~ 현재 : 충북대학교 정보통신공학과(석사과정)

〈관심분야〉 : 빅 데이터, 그래프 스트림, 그래프 마이닝, 데이터베이스 시스템

채 대 성(Daesung Chae)

준회원



- 2001년 2월 : 숭실대학교 전기공학과(공학사)
- 2007년 2월 : 연세대학교 공공정책학과 (정책학석사)
- 2010년 2월 : 국민대학교 정책학과 (정책학박사 수료)
- 2018년 8월 : 충북대학교 빅데이터

학과 (빅데이터학박사 수료)

- 2019년 11월 ~ 현재 : (주)케이에스씨 대표이사
- 〈관심분야〉 : 데이터베이스 시스템, 환경 데이터베이스, 센서 네트워크, 바이오인포매틱스, 빅데이터 등

복 경 수(Kyoungsoo Bok)

종신회원



- 1998년 2월 : 충북대학교 수학과 (이학사)
- 2000년 2월 : 충북대학교 정보통신 공학과(공학석사)
- 2005년 8월 : 충북대학교 정보통신 공학과(공학박사)
- 2005년 3월 ~ 2008년 2월 : 한국 과학기술원 정보전자연구소 Postdoc

- 2008년 3월 ~ 2011년 2월 : 가인정보기술 연구소 차장
 - 2011년 3월 ~ 2019년 8월 : 충북대학교 전자정보대학 정보통신공학부 초빙교수
 - 2011년 9월 ~ 현재 : 원광대학교 SW융합학과 조교수
- 〈관심분야〉 : 데이터베이스 시스템, 이동 객체 데이터베이스, 이동 P2P 네트워크, 소셜 네트워크 서비스, 빅데이터 처리 등

유 재 수(Jaesoo Yoo)

종신회원



- 1995년 2월 : 한국과학기술원 전산학과 (공학박사)
- 1995년 2월 ~ 1996년 8월 : 목포대학교 전산통계학과 전임강사
- 1996년 8월 ~ 현재 : 충북대학교 전자정보대학 정보통신공학부 정교수
- 2009년 8월 ~ 2010년 2월 :

California State University, Fullerton 방문교수

〈관심분야〉 : 데이터베이스 시스템, 멀티미디어 데이터베이스, 센서 네트워크, 바이오인포매틱스, 빅데이터 등