

소셜 네트워크 환경에서 사용자 행위를 고려한 콘텐츠 추천 기법

Contents Recommendation Scheme Considering User Activity in Social Network Environments

고건식*, 김병훈*, 김대윤*, 최민웅*, 임종태**, 복경수**, 유재수**
충북대학교 빅데이터학과*, 충북대학교 정보통신공학과**

Geonsik Ko(kgs@chungbuk.ac.kr)*, Byounghoon Kim(bhkim@chungbuk.ac.kr)*,
Daeyun Kim(kdy0573@chungbuk.ac.kr)*, Minwoong Choi(mwchoi@chungbuk.ac.kr)*,
Jongtae Lim(jtlm@chungbuk.ac.kr)**, Kyoungsoo Bok(ksbok@chungbuk.ac.kr)**,
Jaesoo Yoo(yjs@chungbuk.ac.kr)**

요약

스마트폰의 보급과 온라인 소셜 네트워크 서비스의 발전으로 사용자들은 많은 콘텐츠를 생산하거나 서로 공유한다. 이로 인해 사용자는 자신이 원하지 않는 콘텐츠를 받아보거나 소비함으로써 많은 시간을 소모하게 된다. 이와 같은 문제를 해결하기 위해 소셜 네트워크 사용자에게 적합한 콘텐츠를 추천하기 위한 기법들이 활발하게 연구되고 있다. 본 논문에서는 온라인 소셜 네트워크 사용자에게 협업 필터링을 이용하여 적합한 콘텐츠를 추천하는 기법을 제안한다. 제안하는 기법은 추천의 정확성을 낮추는 사용자의 데이터를 제거하기 위해서 사용자 신뢰도를 고려한다. 사용자의 신뢰도는 온라인 소셜 네트워크의 사용자 행위를 분석해서 도출한다. 사용자의 신뢰도를 다양한 관점에서 평가하기 위해서 기존기법에서 사용하지 않았던 사용자 행위들을 수집해서 사용한다. 다양한 성능평가를 통해 제안하는 기법이 기존 기법보다 우수함을 보인다.

■ 중심어 : | 온라인 소셜 네트워크 | 사용자 행위 | 사용자 신뢰도 | 콘텐츠 추천 |

Abstract

With the development of smartphones and online social networks, users produce a lot of contents and share them with each other. Therefore, users spend time by viewing or receiving the contents they do not want. In order to solve such problems, schemes for recommending useful contents have been actively studied. In this paper, we propose a contents recommendation scheme using collaborative filtering for users on online social networks. The proposed scheme consider a user trust in order to remove user data that lower the accuracy of recommendation. The user trust is derived by analyzing the user activity of online social network. For evaluating the user trust from various points of view, we collect user activities that have not been used in conventional techniques. It is shown through performance evaluation that the proposed scheme outperforms the existing scheme.

■ keyword : | Online Social Network | User Activity | User Trust | Content Recommendation |

* 본 연구는 미래창조과학부 및 정보통신기술진흥센터의 대학CT연구센터육성 지원사업(IITP-2016-H8501-16-1013)과 2015년도 정부(교육부), 2016년도 정부(미래창조 과학부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구사업임 (No.2015R1D1A3A01015962, No.2016R1A2B3007527)

접수일자 : 2016년 10월 24일

수정일자 : 2016년 11월 21일

심사완료일 : 2016년 11월 21일

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

I. 서론

최근 모바일 기기와 인터넷 기술의 발달로 개인의 취미 생활이나 의견을 온라인 공간에서 공유할 수 있는 온라인 소셜 네트워크 서비스가 발전하였다[1]. 온라인 소셜 네트워크는 사용자 간의 자유로운 의사소통, 정보 공유, 그리고 사회적 관계를 생성해주는 온라인 플랫폼을 의미한다. 소셜 네트워크 사용자는 시간, 공간에 상관없이 서비스를 이용할 수 있게 되었고, 인터넷 기술의 발달로 어디서든지 서비스에 빠르게 접속해서 정보와 콘텐츠를 신속하게 생성, 공유할 수 있게 됐다. 콘텐츠는 음악, 영화, 드라마, 뉴스, 게임 등의 사람이 만든 각종 창작물을 의미한다. 대표적인 온라인 소셜 네트워크 서비스로 트위터[2], 페이스북[3], 왓차[4] 등이 존재한다. 트위터와 페이스북은 짧은 글과 함께 사용자의 사진이나 동영상 등을 게시할 수 있는 서비스를 제공한다. 왓차는 영화에 대한 평점, 의견을 공유할 수 있는 서비스를 제공한다.

온라인 소셜 네트워크를 이용하는 사용자가 빠르게 증가하면서 사용자들은 다양한 경로로 수많은 정보를 습득할 수 있게 되었다. 하지만 과도하게 많은 정보에 노출이 되면서 사용자들은 수많은 정보들 중에서 자신에게 적합한 정보만을 습득하기 위한 요구가 점점 커지게 되었다[5-7]. 이런 이유로 사용자의 개인 성향에 따라 필요한 정보를 제공할 수 있는 개인 추천 시스템에 대한 연구가 진행되었다. 개인 추천 시스템에 가장 많이 사용되는 알고리즘은 내용 기반 추천 기법과 협업 필터링 기법이 있다. 내용 기반 추천 기법은 사용자와 콘텐츠 사이에 유사도를 계산해서 연관성이 높은 콘텐츠를 추천하는 기법이다. 이 기법은 사용자와 추천 콘텐츠의 프로필 정보를 모두 알고 있을 때 사용할 수 있다[8]. 협업 필터링은 많은 사용자들로부터 수집한 기호 정보에 따라 사용자들의 관심사를 자동적으로 예측하게 해주는 방법이다. 협업 필터링은 사용자 기반의 협업 필터링과 아이템 기반의 협업 필터링으로 구분된다[9]. 협업 필터링은 사용자가 과거에 경험하지 못한 새로운 정보를 제공할 수 있기 때문에 많은 추천 시스템에 활용되고 있다.

추천 기법에 협업 필터링을 적용하기 위해 사용자의

과거 이력 정보를 활용한다. 해당 사용자와 다른 사용자의 데이터를 이용하는 대표적인 추천 기법이 사용자 기반의 협업 필터링이다. 온라인 소셜 네트워크의 사용자 데이터를 사용하고 데이터를 생성하는 주체인 사용자에 대한 연구가 진행되면서 사용자의 신뢰도를 고려하는 추천기법이 등장하였다. 신뢰도는 사용자를 신뢰할 수 있는지를 나타내는 척도로 사용자의 프로필 정보, 평점, 사용자의 전문성, 사용자간 관계로 표현 가능하다. 사용자의 신뢰도를 고려하는 연구는 개인 신뢰도 모델과 그룹 기반의 신뢰도 모델을 이용하는 연구로 구분된다. 개인 신뢰도 모델은 사용자 개개인의 신뢰도를 계산하는 반면, 그룹 기반의 신뢰도 모델은 네트워크 그룹을 기반으로 그 안에서 유사한 사용자끼리 클러스터링을 한다[10]. 신뢰도 모델을 활용하는 추천기법은 사용자의 신뢰도를 계산하기 위하여 모든 사용자 데이터를 이용한다. 만약 사용자가 악의적으로 콘텐츠에 대해 부정적인 의견을 남기거나 소비하지 않은 콘텐츠를 평가하는 활동을 할 경우, 상대적으로 사용자 신뢰도가 떨어지는 데이터가 추천에 이용된다. 이는 추천의 정확성을 감소시키는 결과를 가져온다. 이러한 문제를 해결하기 위하여 신뢰도가 낮은 사용자의 데이터를 필터링한 후 추천에 이용하는 연구가 등장하였다[11]. 사용자의 신뢰도를 파악하기 위해서는 온라인 소셜 네트워크의 사용자 행위를 이용한다. 기존 연구에서 수행한 사용자 신뢰도 계산은 콘텐츠 사용 횟수를 이용하거나 주변 사람과의 연관성 등을 이용한다. 하지만 기존기법에서 이용한 사용자 행위 외에도 고려할 수 있는 사용자 행위는 다양하다. 따라서 본 논문에서는 기존기법에서 고려한 행위뿐 아니라 다른 사용자 행위를 수집해서 신뢰도를 복합적으로 계산한다.

본 논문에서는 온라인 소셜 네트워크 사용자를 위한 협업 필터링 기반의 콘텐츠 추천 기법을 제안한다. 제안하는 기법은 전체 사용자의 데이터를 사용하지 않기 위해서 신뢰도가 낮은 사용자의 데이터를 필터링한다. 신뢰도를 계산하기 위해서 온라인 환경에서 사용자 행위를 3가지 측면으로 파악하고 분석을 실시한다. 사용자 행위는 현재 서비스 되고 있는 온라인 소셜 네트워크 서비스에서 얻을 수 있는 5개의 행위로 파악한다. 사

용자의 신뢰도를 기반으로 유사 사용자를 선별하고 사용자 기반의 협업 필터링을 수행하여 추천된 콘텐츠의 예측 평점을 계산한다. 최종 결과로 계산된 예측 평점이 높은 순서대로 콘텐츠가 추천된다. 다양한 성능평가를 통해 제안하는 기법의 우수성을 보인다.

본 논문의 구성은 다음과 같다. 제2절에서는 본 논문과 관련된 연구에 대해서 기술한다. 제3절에서는 사용자 행위를 분석해서 신뢰도를 계산하는 기법을 소개한다. 제4절에서는 제안하는 기법의 유효성을 파악하기 위해서 실험으로 검증한다. 마지막으로 제5절에서는 본 논문의 결론과 향후 진행 연구에 대해 기술한다.

II. 관련 연구

콘텐츠 추천 시스템에 대표적으로 사용하는 콘텐츠 추천 알고리즘은 내용기반 추천 알고리즘과 협업 필터링 알고리즘이다. 내용기반 추천 알고리즘은 아이템이나 사용자를 분석해서 비슷한 아이템을 추천한다. 아이템과 사용자 사이의 관계를 분석하는 것이 아니라 콘텐츠 자체를 분석하기 때문에 사용자의 행동과 관계없이 추천을 할 수 있다. 따라서 새로운 콘텐츠가 추천 시스템에 들어오더라도 콘텐츠에 대한 분석만으로도 추천이 될 수 있다. 협업 필터링 알고리즘은 아이템이나 유저의 유사도를 모델링하고 이를 측정해서 추천한다. 그림 1은 협업 필터링 알고리즘을 나타낸다. 사용자 기반 협업 필터링은 사용자의 행위를 분석하고 사용자 사이의 유사도를 측정해서 유사한 사용자를 찾고 유사 사용자가 사용한 콘텐츠를 추천한다. 아이템 기반 협업 필터링은 사용자가 사용한 아이템과 유사도가 높은 다른 아이템을 추천한다.

신뢰도가 낮은 사용자를 필터링하기 위해서 신뢰도를 계산하는 여러 추천 기법이 존재한다. P. Moradi 등은 추천 시스템의 정확도를 향상시키기 위해서 신뢰도를 기반으로 유사한 사용자를 선별하는 기법을 제안하였다[12]. 이 기법은 사용자가 콘텐츠에 부여한 평점을 이용해서 피어슨 유사도를 계산한다. 계산된 유사도를 기반으로 1차 신뢰도 네트워크를 구성한다. 다음 단계에서 사용자에 대한 긍정, 부정 요인을 계산한 값이 임

계치를 만족하지 못하면 1차 신뢰도 네트워크에서 사용자를 제거하고 2차 신뢰도 네트워크를 구성한다. 완성된 신뢰도 네트워크를 이용해서 추천 시스템을 구성한다.

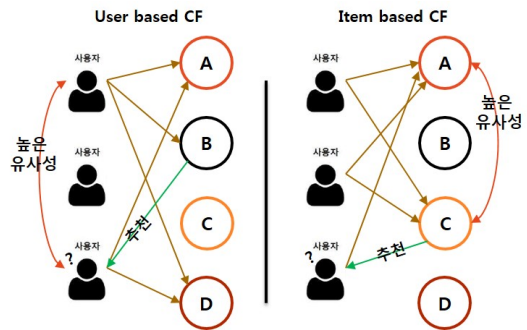


그림 1. 협업 필터링 알고리즘

H. Dimah 등은 사용자의 친구가 게재한 트윗 내용을 이용한 감정분석과 팔로우·팔로잉 관계를 이용한 추천 기법을 제안하였다[13]. 이 기법은 사용자와 사용자의 친구가 서로 의사소통을 통해 영향을 받는다는 사실을 이용한다. 따라서 사용자의 친구가 추천할 콘텐츠에 대해서 게재한 트윗을 이용한다. POS Tagging 기법을 이용해서 단어를 분류하고 분류된 단어를 이용해서 감정 분석을 실시한다. 감정 분석을 통해 친구들이 추천 콘텐츠에 대해 평가한 예측 점수를 계산한다. 사용자의 팔로잉, 팔로워 수를 이용해서 사용자의 신뢰도를 계산한다. 이때 계산된 신뢰도는 해당 사용자를 신뢰하는 정도가 아니라 사용자의 친구를 얼마나 신뢰하는지를 의미한다. 즉, 계산된 신뢰도가 클수록 사용자는 친구들의 영향을 많이 받는다는 의미이다. 계산된 예측 점수와 사용자의 신뢰도를 이용해서 최종적으로 추천 콘텐츠에 대한 예상 점수를 계산한다.

W. Hwang 등은 콘텐츠 사용자 중에서 전문가를 선별하는 기법을 제안하였다[14]. 이 기법은 사용자가 사용한 콘텐츠가 속한 장르를 얼마나 사용 하였는가를 기반으로 전문가를 선별한다. 선별된 전문가는 장르별 전문가로 구분된다. 전문가와 해당 사용자 사이의 유사도를 계산해서 최종적으로 추천 콘텐츠에 대한 예상 점수를 계산한다.

기존 논문들 중에서 [11]에서는 사용자 사이의 유사도에 영향을 주는 긍정, 부정 요인을 계산하는 식만 제안하고 신뢰도 사용자를 선별하는 기준이 되는 사용자 행위 등은 고려하지 않고 있다. 따라서 사용자의 신뢰도를 판단하기 위한 사용자 행위에 대한 연구가 필요하다. [12] 과 [13]에서 제안한 기법들은 사용자의 신뢰도를 판단하기 위해서 사용자 행위를 고려하고 있다. 하지만 각각의 기법은 1개 또는 2개의 사용자 행위를 이용할 뿐 복합적으로 분석하고 있지 않다. 또한 현재 서비스되는 온라인 소셜 네트워크에서 얻을 수 있는 사용자 행위들을 모두 포함하고 있지 않다. 따라서 본 논문에서는 기존기법에서 고려하는 사용자 행위 외에 추가로 다른 성격의 사용자 행위를 수집해서 소셜 행위 분석[15], 콘텐츠 이용 분석, 소셜 관계 분석을 통해 사용자의 신뢰도를 계산한다.

III. 제안하는 콘텐츠 추천 기법

1. 제안하는 기법의 주요 특성

본 논문에서는 온라인 소셜 네트워크 사용자의 데이터를 활용해서 신뢰도가 떨어지는 사용자를 필터링하고 협업필터링을 통해 콘텐츠를 추천하는 기법을 제안한다. 현재 서비스되고 있는 온라인 소셜 네트워크 환경에서 수집할 수 있는 사용자 행위를 이용한다. 그리고 추천 정확성을 높이기 위해서 사용자 신뢰도와 콘텐츠 신뢰도를 이용한다. 소셜 행위 분석, 콘텐츠 이용 분석, 소셜 관계 분석을 통해서 사용자 신뢰도를 계산하고, 사용자 전문성 분석과 암시적 행위 분석을 통해서 콘텐츠 신뢰도를 계산한다.

[그림 2]는 제안하는 기법의 전체적인 처리 절차를 나타내고 있다. 제안하는 기법에서는 온라인 소셜 네트워크를 이용하는 사용자의 행위 데이터를 수집한다. 수집된 사용자 행위는 기존 기법에서 고려하지 않았던 사용자 행위를 포함하면서 여러 측면으로 사용자 신뢰도를 평가한다. 수집된 데이터를 이용해서 신뢰할 수 있는 사용자를 선별하기 위해서 사용자 신뢰도를 계산한다. 사용자 행위를 바탕으로 소셜 행위 분석, 콘텐츠 이

용 분석, 소셜 관계 분석 등을 수행한다. 계산된 사용자 신뢰도를 바탕으로 신뢰할 수 없는 사용자를 필터링한다. 필터링하고 난 뒤 남은 신뢰할 수 있는 사용자와 해당 사용자의 유사성을 계산하고 사용자 기반 협업 필터링을 통해서 추천할 콘텐츠를 선별한다. 그 다음 추천할 콘텐츠의 예측 평점을 계산한 후, 계산된 예측 평점의 순서대로 콘텐츠를 정렬해서 점수가 높은 순서대로 콘텐츠 추천을 실시한다.



그림 2. 제안하는 기법의 처리 절차

2. 사용자 행위

추천의 정확성을 향상시키기 위해서 신뢰도가 낮은 데이터를 필터링해야 한다. 필터링을 위해 사용자의 신뢰도를 우선 계산해야 한다. 사용자 신뢰도는 사용자 행위 분석을 통해서 도출하는데, 기존 기법에서는 온라인 소셜 네트워크 서비스 환경에서 수집할 수 있는 사용자 행위의 일부만 반영하고 있다. 따라서 기존기법에서 고려하지 않고 있는 행위를 포함한 5가지의 사용자의 행위를 수집해서 분석에 이용한다. 수집된 사용자 행위를 바탕으로 소셜 행위, 콘텐츠 이용, 소셜 관계 등의 3가지 측면으로 구분해서 신뢰도 분석을 실시한다.

[표 1]은 본 논문에서 신뢰도 계산을 위한 사용자 행

위를 나타낸다. 수집된 행위들은 각 코멘트가 받은 좋아요 수, 사용자가 쓴 코멘트 수, 사용 콘텐츠 수, 팔로잉 수, 팔로워 수 등 총 5가지이다. [표 1]에서 제안하는 사용자 행위는 온라인 소셜 네트워크 서비스에서 사용자들이 콘텐츠를 소비하고 공유하는 활동을 통해서 자연스럽게 남겨진 활동들이기 때문에 수집하기 용이하다.

표 1. 사용자 행위와 측도 표현

사용자 행위	좋아요 수	코멘트 수	사용 콘텐츠 수	팔로잉 수	팔로워 수
측도	A	B	C	F_{ing}	F_{wer}

A 는 온라인 소셜 네트워크 사용자가 콘텐츠에 대한 코멘트를 남겼을 때 다른 사용자들이 코멘트에 대해 주는 '좋아요 수'를 의미한다. 좋아요 수를 많이 받은 사용자를 신뢰 수 있다고 가정한다. [그림 3]는 해당 사용자의 좋아요 수를 구하는 과정을 나타낸다.

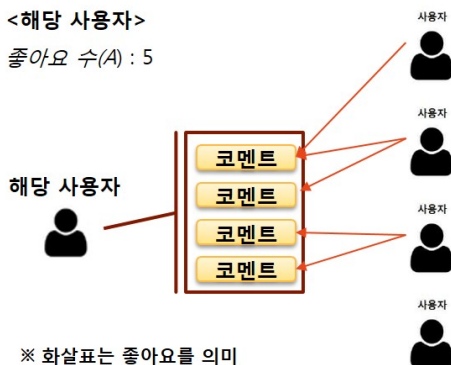


그림 3. 해당 사용자의 좋아요 수(A)

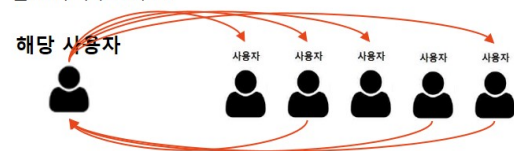
B 는 사용자가 이용한 콘텐츠에 대해서 남긴 '코멘트 수'를 의미한다. C 는 사용자가 이용한 '모든 콘텐츠 수'를 의미한다. F_{ing} 은 해당 사용자의 '팔로잉(Following) 수'를 의미한다. 팔로잉이란, 해당 사용자가 다른 사용자의 정보를 받아보겠다는 행위로 해당 사용자의 팔로잉 수가 높으면 다른 사용자의 정보를 많이 받아 보겠

다는 의미이기 때문에 해당 사용자의 의존성을 나타내기도 한다. F_{wer} 은 해당 사용자의 '팔로워(Follower) 수'를 의미한다. 팔로워는 해당 사용자를 팔로잉 한 다른 사용자를 의미한다. 즉, 다른 사용자들로부터 받은 모든 팔로잉의 합으로 표현할 수 있다. [그림 4]는 사용자의 팔로잉과 팔로워 수를 나타낸다.

<해당 사용자>

팔로잉 수(F_{ing}): 4

팔로워 수(F_{wer}): 3



※ 화살표는 팔로잉을 의미

그림 4. 해당 사용자의 팔로잉, 팔로워 수

3. 사용자 신뢰도 계산

3.1 소셜 행위 분석

온라인 소셜 네트워크 공간에서의 해당 사용자의 소셜 행위를 분석해서 소셜 활동에 대한 평가를 한다. 소셜 행위 분석에는 '좋아요 수' A 와 '코멘트 수' B 를 이용한다. 사용자의 소셜 행위인 좋아요 수와 코멘트 수에 따라 사용자의 신뢰도 점수가 결정된다. 소셜행위를 통해 해당 사용자의 의견이 다른 사용자에게 어떻게 받아들여졌는지, 해당 사용자가 받은 평판은 어떤지가 평가에 포함된다. 사용자 신뢰도 점수 K 는 수식 (1)로 표현한다. 사용자 신뢰도 점수 K 는 사용자가 받은 좋아요 수와 코멘트 수의 합에서 사용자가 받은 좋아요 수가 차지하는 비율을 나타낸다. 신뢰도 점수 K 는 0부터 1사이의 값을 가지게 된다. 사용자가 남긴 코멘트의 수가 적을수록, 받은 좋아요 수가 많을수록 신뢰도 점수 K 의 값은 1에 근사한 값을 가진다. 계산된 신뢰도 값이 0에 가깝다면 사용자의 소셜 행위가 다른 사람의 공감을 얻지 못했다는 의미를 뜻하기 때문에 해당 사용자를 신뢰하지 못한다는 의미이다.

$$K = A / (A + B) \quad (1)$$

3.2 콘텐츠 이용 분석

사용자의 콘텐츠 이용 횟수를 이용한 분석을 위해서 사용자가 이용한 콘텐츠 수를 바탕으로 사용자 신뢰도 점수 C_o 를 계산한다. 사용자 신뢰도 점수 C_o 는 수식 (2)로 표현한다. C_{all} 은 제안하는 기법에서는 소셜 네트워크에서 관련된 콘텐츠를 이용한 횟수가 적은 사용자보다는 상대적으로 콘텐츠를 이용한 횟수가 많은 사용자가 더 높은 신뢰도를 가진다고 가정하였다. 따라서 사용자가 이용한 콘텐츠의 수가 높을수록 더 높은 신뢰도 점수를 가진다.

$$C_o = C / C_{all} \quad (2)$$

3.3 소셜 관계 분석

사용자의 소셜 관계를 분석하기 위해서 온라인 소셜 네트워크 사용자가 가지고 있는 ‘팔로잉 수’와 ‘팔로워 수’를 바탕으로 사용자 신뢰도 점수 F 를 계산한다. 팔로워 수가 높다는 것은 나의 정보를 수신하길 원하는 다른 사용자의 수가 많다는 것을 의미한다. 즉 내가 다른 사용자에게 영향을 주고 있음을 의미한다. 팔로잉 수는 내가 다른 사용자의 정보를 수신하는 것을 의미한다. 즉 팔로잉은 내가 다른 사용자로부터 받는 영향력을 의미한다. 따라서 상대적으로 팔로잉 수가 많은 것보다는 팔로워 수가 많을 때 그 사용자는 온라인 소셜 네트워크 상에서 외부 영향력이 크다고 가정한다. 외부 영향력이란 사용자가 다른 사용자에게 주는 영향력의 크기를 의미한다. 외부 영향력이 클수록 다른 사용자에게 미치는 영향력이 크기 때문에 신뢰도 있는 사용자로 판단한다. 사용자 신뢰도 점수 F 는 수식 (3)로 표현한다. 사용자 신뢰도 점수 F 는 사용자의 팔로잉 수와 팔로워 수의 합에서 팔로워 수가 차지하는 비율을 나타낸다. 팔로워 수가 많을수록, 팔로잉 수가 적을수록 사용자 신뢰도 점수 F 의 값은 1에 가깝다.

$$F = F_{wer} / (F_{ing} + F_{wer}) \quad (3)$$

계산된 3개의 사용자 신뢰도 점수 K , C , F 를 종합해서 최종 사용자 신뢰도 점수 T 를 계산한다. T 는 수식 (4)로 표현한다. 수식 (4)의 α, β, γ 는 각각 계산된 신뢰도 점수에 대해 부여한 가중치를 의미한다. α, β, γ 의 합은 1이다. 최종 계산된 사용자 신뢰도 T 와 사용자가 정한 임계치 θ 를 비교해서 임계치보다 작은 신뢰도를 가진 사용자를 필터링한다. 필터링 후 남은 사용자들을 신뢰할 수 있는 사용자로 선별한다. 선별된 사용자는 신뢰할 수 있는 사용자 그룹으로 구분된다. [그림 5]는 사용자의 신뢰도를 계산하고 신뢰도를 바탕으로 사용자를 필터링하는 알고리즘을 나타낸다.

$$T = \alpha K + \beta C_o + \gamma F \quad (\alpha + \beta + \gamma = 1) \quad (4)$$

User_Trust_Evaluator_Filtering()	
01	{ $U_i(K, C, F)$
02	$K = \text{social_Analysis}(A, B);$
03	$C_o = \text{content_Analysis}(C);$
04	$F = \text{relation_Analysis}(F_{ing}, F_{wer});$
05	$\text{sum}(T) = \alpha K + \beta C + \gamma F;$
06	if ($\text{sum}(T) > \theta$) then
07	{
08	$\text{trust user} = U_i;$
09	return ;
10	}
11	}

그림 5. 제안하는 사용자 필터링 알고리즘

4. 협업 필터링

사용자에게 추천할 콘텐츠를 결정하기 위해서 협업 필터링을 실시한다. 협업 필터링을 실시하기 위해서는 해당 사용자와 유사한 성향을 가진 사용자의 데이터가 필요하다. 따라서 앞서 선별된 신뢰할 수 있는 사용자와 해당 사용자의 유사도를 계산해서 찾은 유사 사용자의 데이터를 활용한다. 사용자의 콘텐츠 과거 평점 이력을 바탕으로 피어슨 유사도를 계산한다. 피어슨 유사도는 수식 (5)로 표현한다. u 는 해당 사용자, e 는 선별된 신뢰할 수 있는 사용자를 의미한다. i 는 과거 평점을 준 콘텐츠 객체, I 는 전체 콘텐츠의 집합을 의미한다. 유사도 계산을 통해 해당 사용자와 유사한 사용자를 선별하게 되고, 유사 사용자를 이용해서 사용자 기반 협업 필터링을 실시한다. 협업 필터링을 통해서 해당 사

용자가 이용하지 않은 추천 콘텐츠 리스트를 결정한다.

$$S(u, e) = \frac{\sum_{i \in I} (r_i(u) - \bar{r}(u))(r_i(e) - \bar{r}(e))}{\sqrt{\sum_{i \in I} (r_i(u) - \bar{r}(u))^2} \sqrt{\sum_{i \in I} (r_i(e) - \bar{r}(e))^2}} \quad (5)$$

5. 콘텐츠 추천

추천 콘텐츠를 해당 사용자에게 순서대로 추천해 주기 위해서 콘텐츠에 대한 예측 평점을 계산한다. 계산된 예측 평점은 해당 사용자가 추천된 콘텐츠를 사용하고 준 평점으로 예측한다. 예측 평점을 기준으로 추천 콘텐츠를 정렬하고 예측 평점이 높은 순서대로 콘텐츠를 추천한다. 식(6)은 예측 평점을 계산하는 수식이다. u 는 해당 사용자, i 는 추천 콘텐츠를 의미한다. E 는 유사 사용자 그룹이고 e 는 유사 사용자 객체를 의미한다. 식(5)에서 계산된 유사도를 이용해서 추천 콘텐츠의 예측 평점을 계산한다.

$$P(u, i) = \bar{r}_u + \frac{\sum_{e \in E} S(u, e)(r_{(e, i)} - \bar{r})}{\sum_{e \in E} S(u, e)} \quad (6)$$

IV. 성능 평가

본 절에서는 제안하는 기법의 추천의 정확도와 정밀도를 평가한다. 성능평가는 본 논문에서 제안하는 기법과 [8]에서 제안된 기본적인 아이템 기반 협업 필터링과 사용자 기반 협업 필터링 기법을 비교 평가한다. 실험환경은 Intel core i3-2300 CPU 3.10GHz 프로세서와 6.00GB RAM, Windows7 환경을 가진 데스크톱에서 Java로 구현하였다. 추천 콘텐츠는 영화를 대상으로 하였으며 사용자 행위 데이터는 왓차 어플을 이용하는 사용자에서 무작위로 100명의 사용자를 선택해서 수집하였다. 영화 평점 데이터는 미네소타 대학의 GroupLens에서 제공하는 MovieLens 데이터를 활용하였다. 수집된 데이터의 부족한 정보를 보완하기 위해 사용자 행위 데이터와 영화 평점 데이터를 유저를 기준으로 맵핑하

여 하나의 데이터로 처리하였다. 성능평가에 사용된 사용자 평점의 총 개수는 10874개, 사용된 전체 영화의 수는 855개이다. 수집된 실험 데이터를 이용해서 추천 기법을 통해 예측한 평점 값과 실제 사용자가 콘텐츠에 준 평점 값을 비교한다. 두 값을 비교해서 추천의 정확성이 향상되었는지를 판단한다. 추천의 정확성을 판단하는 기준은 *Mean Absolute Error(MAE)*, *Root Mean Square Error(RMSE)* 측도를 이용한다. P 는 예측평점을 의미하고 R 은 실제로 준 평점을 의미한다. MAE 는 실제 값과 예측한 값의 오차의 평균을 의미한다. MAE 가 작을수록 정확도가 좋은 것이다. $RMSE$ 는 평균 제곱 오차의 제곱근을 의미한다. $RMSE$ 가 작을수록 정밀도가 좋은 것이다.

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_{(u, i)} - R_{(u, i)}| \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_{(u, i)} - R_{(u, i)})^2} \quad (8)$$

실험을 위해 우선 사용자 신뢰도를 계산해야한다. 그런데 신뢰도를 계산하기 전에 신뢰도 함수에 들어가는 가중치를 결정해야한다. 신뢰도를 구하기 위해서는 소셜 행위분석, 콘텐츠 이용분석, 소셜 관계 분석을 통해 각각의 신뢰도를 구한다. 가중치에 의미를 판별하기 위해 서로 다른 가중치를 부여한 제안 추천 모델이 나타내는 사용자 신뢰도 T 를 구한다. [표 2]는 3개의 분석으로 각각 구한 신뢰도 점수 K , C_o , F 에 서로 다른 가중치를 적용한 case1-5을 보여준다. case1은 소셜 행위 분석과 소셜 관계 분석에 대해 동일한 가중치 비중을 주기 위해서 각각 0.4의 가중치를 주고 콘텐츠 이용 분석에는 0.2의 가중치를 주었다. case2와 case3은 소셜 행위 분석, 소셜 관계 분석, 콘텐츠 이용 분석의 순서대로 가중치의 비중을 다르게 주었다. case4는 콘텐츠 이용 분석과 소셜 관계 분석에 대해 동일한 가중치 비중을 주기 위해서 각각 0.3의 비중을 주고 소셜 행위 분석에는 0.4의 가중치를 주었다. case5는 소셜 관계 분석, 소셜 행위 분석, 콘텐츠 이용 분석 순서대로 가중치의 비중을 다르게 주었다.

표 2. case1-5 의 가중치

	case1	case2	case3	case4	case5
α	0.4	0.5	0.6	0.4	0.3
β	0.2	0.2	0.1	0.3	0.2
γ	0.4	0.3	0.3	0.3	0.5

[그림 6]은 Case1-5의 5가지의 서로 다른 신뢰도 가중치를 주었을 때, 사용자 100명이 가지는 신뢰도 값 T 을 보여준다. 그래프에서 나타나는 수치를 서로 비교하였을 때, case3일 때의 T 가 전반적으로 높은 값을 가지는 것을 확인할 수 있다.

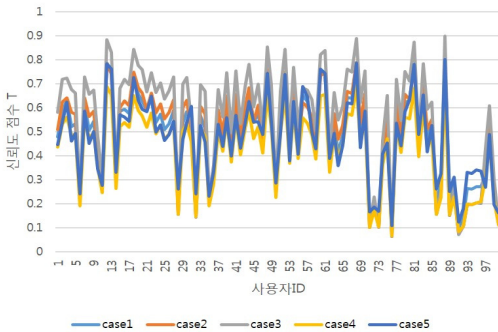


그림 6. 가중치에 따른 사용자 신뢰도

[표 3]은 서로 다른 비율의 가중치를 구했을 때 100명의 사용자가 가지는 신뢰도 점수의 분산을 나타낸다. [표 3]에서 나타내는 분산의 크기가 클수록 점수의 분포가 평균값에 몰려 있지 않고 퍼져있음을 의미한다. 5가지의 case들 중에서 case3인 경우에 신뢰도 점수가 가지는 분산의 값이 가장 크다. 분산이 클수록 신뢰도 점수 값이 넓게 퍼져 있고 값이 넓게 퍼져 있을수록 사용자가 가질 수 있는 신뢰도 값의 폭은 커지게 된다. 사용자가 가지는 신뢰도 값의 폭이 넓을 경우에 신뢰도를 이용해서 사용자를 필터링할 때 사용자들을 구분하기가 더 용이하다. 결과적으로 다른 case의 경우에는 0~0.8 사이의 값을 가지지만 case3의 경우에 신뢰도 값의 범위가 0~0.9 사이로 가장 넓다. 따라서 case3에서 사용된 가중치인 $\alpha = 0.6$, $\beta = 0.1$, $\gamma = 0.3$ 의 값을 성능 평가에 적용한다.

표 3. 가중치에 따른 사용자 신뢰도의 분산

가중치	case1	case2	case3	case4	case5
분산	0.0321	0.0389	0.0528	0.0273	0.0275

신뢰도가 낮은 사용자를 필터링하기 위해서 계산된 신뢰도와 비교할 임계치 값 θ 를 결정해야 한다. θ 값을 다양하게 주면서 각 경우의 MAE 값을 계산한다. MAE를 계산하는 이유는 추천의 정확도를 파악해서 정확도가 가장 높을 때의 θ 값을 찾아서 이용하기 위함이다. [그림 7]은 θ 값에 따라서 변화하는 제안 기법의 MAE를 보여주고 있다. 결과적으로 0.7일 때 MAE의 값이 0.735의 MAE 값을 가진다. 따라서 추천 정확성이 가장 좋은 $\theta = 0.7$ 을 성능평가에 사용한다.

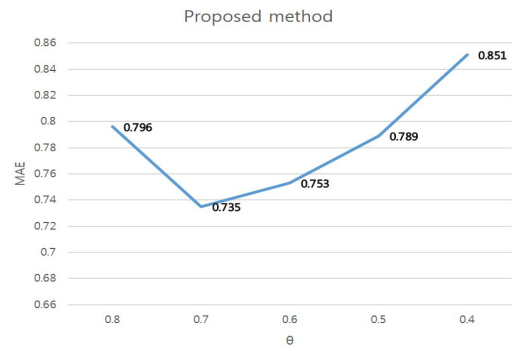


그림 7. θ 값에 따른 제안하는 기법의 MAE 변화

[표 4]은 제안하는 알고리즘을 이용해서 임의의 사용자를 해당 사용자로 선택하고 시뮬레이션을 통해서 나온 5개의 추천 영화 목록이다. 예상 평점이 높은 순서대로 추천 영화를 선별한다. 추천 영화의 예상 평점과 영화를 시청한 후 사용자가 준 실제 평점을 비교해서 추천의 정확성과 정밀도를 평가한다.

표 4. 해당 사용자에게 추천할 영화 목록

User ID	Order	Movie ID
100	1	50
100	2	61
100	3	4
100	4	783
100	5	628

본 논문에서 제안하는 기법의 정확도와 정밀도를 평가하기 위해서 동일한 데이터와 동일한 환경으로 [8]의 아이템 기반 협업 필터링과 사용자 기반 협업 필터링, [12]의 P.Moradi의 방법(P.Moradi's method)과 제안하는 기법을 비교 성능 평가를 수행하였다. [그림 8]와 [그림 9]은 각각 제안하는 기법과 [8]과 [12]의 MAE와 RMSE를 나타낸다. [그림 8]에서는 제안하는 기법의 MAE가 아이템 기반의 협업 필터링을 실시해서 구한 MAE보다 11% 낮고 사용자 기반의 협업 필터링을 실시해서 구한 MAE보다 12% 낮다. 또한 [12]를 실시해서 구한 MAE보다 6% 낮은 것을 확인할 수 있다. [그림 9]에서는 제안하는 기법의 RMSE가 아이템 기반의 협업 필터링을 실시해서 구한 RMSE보다 7% 낮고 사용자 기반의 협업 필터링을 실시해서 구한 RMSE보다 6% 낮은 수치를 보여준다. 또한 [12]를 실시해서 구한 RMSE보다 3% 낮다. 성능평가 결과를 보면 기존기법보다 제안하는 기법의 추천 정확도와 정밀도가 좋은 것을 확인할 수 있다.

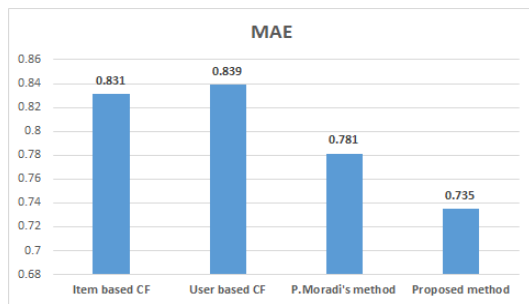


그림 8. 제안 기법과 기존 기법의 MAE 비교

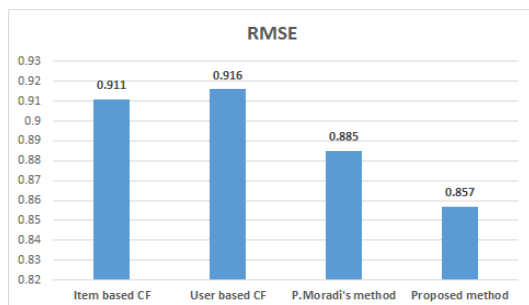


그림 9. 제안 기법과 기존 기법의 RMSE 비교

V. 결 론

본 논문에서는 사용자 행위를 고려한 사용자 신뢰도 기반의 효율적인 콘텐츠 추천 기법을 제안하였다. 기존 논문에서 다양한 사용자 행위를 이용해서 신뢰성을 계산하지 않고 있기 때문에 현재 서비스되고 있는 온라인 소셜 네트워크에서 5가지의 사용자 행위를 수집해서 신뢰도를 계산한다. 사용자 신뢰도는 소셜 행위 분석, 콘텐츠 이용 분석, 소셜 관계 분석 등 3개의 분석을 이용해서 계산한다. 실험을 통하여 제안하는 기법은 기존 기법에 비해 추천의 정확도와 정밀도가 향상됨을 확인할 수 있었다. 향후 연구로는 사용자의 전문성을 바탕으로 콘텐츠의 신뢰도를 계산해서 추천 콘텐츠를 결정하는 연구와 사용자의 성향을 고려한 연구를 진행할 것이다.

참 고 문 헌

- [1] 나종연, “사용확산모형을 적용한 소비자의 온라인 소셜 네트워크 활용에 대한 연구,” 소비자학연구, 제21권, 제2호, pp.443-472, 2010.
- [2] <https://www.twitter.com/>
- [3] <https://ko-kr.facebook.com/>
- [4] <https://www.watcha.net/>
- [5] 고상민, 황보환, 지용구, “소셜네트워크서비스와 온라인 사회적 자본- 한국과 중국 사례를 중심으로- 페이스북 이용을 중심으로,” 한국전자거래학회지, 제15권, 제1호, pp.103-118, 2010.
- [6] 김유정, “소셜네트워크서비스에 대한 이용과 충족 연구,” 미디어, 젠더 & 문화, 제20호, pp.71-105, 2011.
- [7] 송창우, 김종훈, 정경용, 류중경, 이정현, “시맨틱 웹에서 개인화 프로파일을 이용한 콘텐츠 추천 검색 시스템,” 한국콘텐츠학회논문지, 제8권, 제1호, pp.318-327, 2008.
- [8] L. Iaquinta, M. de Gemmis, P. Lops, G. Semeraro, M. Filannino, and P. Molino, “Introducing serendipity in a content-based recommender

- system,” Proc. International Conference on Hybrid Intelligent Systems, pp.168-173, 2008.
- [9] M. HAHLER, “recommenderlab: A Framework for Developing and Testing Recommendation Algorithms,” 2011.
- [10] A. Javari and M. Jalili, “Cluster-based collaborative filtering for sign prediction in social networks with positive and negative links,” ACM Transactions on Intelligent Systems and Technology, Vol.5, No.2, p.24, 2014
- [11] S. Gong, “A collaborative filtering recommendation algorithm based on user clustering and item clustering,” Journal of Software, Vol.5, No.7, pp.745-752, 2010.
- [12] P. Moradi and S. Ahmadian, “A reliability-based recommendation method to improve trust-aware recommender systems,” Expert Systems with Applications, Vol.42, No.21, pp.7386-7398, 2015.
- [13] D. H. Alahmadi and X. Zeng, “ISTS: Implicit social trust and sentiment based approach to recommender systems,” Expert Systems with Applications, Vol.42, No.22, pp.8840-8849, 2015.
- [14] W. Hwang, H. Lee, S. Kim, Y. Won, and M. Lee, “Efficient recommendation methods using category experts for a large dataset,” Information Fusion, Vol.28, pp.75-82, 2016.
- [15] 노연우, 김대윤, 한지은, 육미선, 임종태, 복경수, 유재수, “소셜 네트워크에서 사용자의 영향력을 고려한 핫 토픽 예측 기법,” 한국콘텐츠학회논문지, 제15권, 제8호, pp.24-36, 2015.

저 자 소 개

고 건 식(Geonsik Ko)

준회원



- 2015년 2월 : 건국대학교 응용통계학과(경제학사)
- 2015년 3월 ~ 현재 : 충북대학교 빅데이터 협동과정(석사과정)

<관심분야> : 빅데이터, 데이터베이스 시스템, 소셜 네트워크 서비스 등

김 병 훈(Byounghoon Kim)

준회원



- 2015년 2월 : 청주대학교 컴퓨터정보공학과(공학사)
- 2015년 3월 ~ 현재 : 충북대학교 빅데이터 협동과정(석사과정)

<관심분야> : 빅데이터, RDF, 분산 질의 처리, 데이터베이스 시스템 등

김 대 윤(Daeyun Kim)

준회원



- 2015년 2월 : 청주대학교 경영학과/컴퓨터공학과(공학사)
- 2015년 3월 ~ 현재 : 충북대학교 빅데이터 협동과정(석사과정)

<관심분야> : 데이터베이스 시스템, 복합이벤트처리, 빅데이터 등

최 민 웅(Minwoong Choi)

준회원



- 2015년 2월 : 신라대학교 컴퓨터 공학과, 경영학과(경영학 학사, 컴퓨터 공학사)
- 2015년 3월 ~ 현재 : 충북대학교 빅데이터 협동과정(석사과정)

<관심분야> : 데이터마이닝, 빅데이터, 사물인터넷, 경영전략 등

임 중 태(Jongtae Lim)

정회원



- 2009년 2월 : 충북대학교 정보통신공학과(공학사)
- 2011년 2월 : 충북대학교 정보통신공학과(공학석사)
- 2011년 3월 ~ 현재 : 충북대학교 정보통신공학과(공학박사)

<관심분야> : 데이터베이스 시스템, 시공간 데이터베이스, 위치기반 서비스, 모바일 P2P 네트워크, 빅데이터 등

유 재 수(Jaesoo Yoo)

중신회원



- 1989년 2월 : 전북대학교 컴퓨터 공학과(공학사)
- 1991년 2월 : KAIST 전산학과(공학석사)
- 1995년 2월 : KAIST 전산학과(공학박사)

▪ 1995년 3월 ~ 1996년 8월 : 목포대학교 전산통계학과 (전임강사)

▪ 1996년 8월 ~ 현재 : 충북대학교 정보통신공학부 및 컴퓨터정보통신연구소 교수

▪ 2009년 3월 ~ 2010년 2월 : 캘리포니아주립대학교 방문교수

<관심분야> : 데이터베이스 시스템, 빅데이터, 센서네트워크 및 RFID, 소셜 네트워크 서비스, 분산 객체컴퓨팅, 바이오인포매틱스 등

북 경 수(Kyungsoo Bok)

중신회원



- 1998년 2월 : 충북대학교 수학과(이학사)
- 2000년 2월 : 충북대학교 정보통신공학과(공학석사)
- 2005년 2월 : 충북대학교 정보통신공학과(공학박사)

▪ 2005년 3월 ~ 2008년 2월 : 한국과학기술원 전산학과 Postdoc

▪ 2008년 3월 ~ 2011년 2월 : (주)가인정보기술 연구소

▪ 2011년 3월 ~ 현재 : 충북대학교 정보통신공학과 초빙교수

<관심분야> : 데이터베이스 시스템, 이동객체 데이터베이스, 소셜 네트워크 서비스, 빅데이터 등