ISSN 2288-4866 (Print)
ISSN 2288-4882 (Online)
http://www.jiisonline.org
J Intell Inform Syst 2015 June: 21(2): 1~18 http://dx.doi.org/10.13088/jiis.2015.21.2.01

사용자 리뷰 마이닝을 결합한 협업 필터링 시스템: 스마트폰 앱 추천에의 응용*

전병국
국민대학교 비즈니스IT전문대학원 석사과정
(boungkug@kookmin.ac.kr)

안현철
국민대학교 비즈니스IT전문대학원 부교수
(hcahn@kookmin..ac.kr)

협업 필터링은 학계나 산업계에서 우수한 성능으로 인해 많이 사용되는 추천기법이지만, 정량적 정보인 사용자들의 평가점수에만 국한하여 추천결과를 생성하므로 간혹 정확도가 떨어지는 문제가 발생한다. 이에 새로운 정보를 추가로 고려하여, 협업 필터링의 성능을 개선하려는 연구들이 지금까지 다양하게 시도되어 왔다. 본 연구는 최근 Web 2.0 시대의 도래로 인해 사용자들이 구입한 상품에 대한 솔직한 의견을 인터넷 상에 자유롭게 표현한다는 점에 착안하여, 사용자가 직접 작성한 리뷰를 참고하여 협업 필터링의 성능을 개선하는 새로운 추천 알고리즘을 제안하고, 이를 스마트폰 앱 추천 시스템에 적용하였다. 정성 정보인 사용자 리뷰를 정량화하기 위해 본 연구에서는 텍스트 마이닝을 활용하였다. 구체적으로 본 연구의 추천시스템은 사용자간 유사도를 산출할 때, 사용자 리뷰의 유사도를 추가로 반영하여 보다 정밀하게 사용자간 유사도를 산출할 수 있도록 하였다. 이 때, 사용자 리뷰의 유사도를 산출하는 접근법으로 중복 사용된 색인어의 빈도로 산출하는 방안과 TF-IDF(Term Frequency - Inverse Document Frequency) 가중치 합으로 산출하는 2가지 방안을 제시한 뒤 그 성능을 비교해 보았다. 실험결과, 제안 알고리즘을 통한 추천, 즉 사용자 리뷰의 유사도를 추가로 반영하는 알고리즘이 평점만을 고려하는 전통적인 협업 필터링과 비교해 더 우수한 예측정확도를 나타냄을 확인할 수 있었다. 아울러, 중복 사용 단어의 TF-IDF 가중치의 합을 고려했을 때, 단순히 중복 사용 단어의 빈도만을 고려했을 때 보다 조금 더 나은 예측정확도를 얻을 수 있음도 함께 확인할 수 있었다.

주제어 : 추천시스템, 협업 필터링, 텍스트 마이닝, TF-IDF, 앱스토어

논문접수일 : 2015년 5월 20일 논문수정일 : 2015년 6월 16일 게재확정일 : 2015년 6월 16일
투고유형 : 학술대회 우수논문 교신저자 : 안현철

1. 서론

협업 필터링(Collaborative filtering, CF) 추천 시스템은 추천 대상이 되는 고객과 비슷한 패턴을 가진 고객들(이웃)을 식별하여, 이들이 과거에 선호했거나 구매했던 상품들 중 대상 고객이 아직 경험하지 않은 상품을 추천하는 시스템이다. 이러한 협업 필터링 기반의 추천 시스템을 구현하기 위해서는 추천 대상이 되는 고객과 유사한 고객이 누구인지를 얼마나 정확하게 식별하는가가 중요한데, 전통적인 협업 필터링에서 고객 간 유사도는 상품에 대한 평점 혹은 구매 여부와 같은 정량적(quantitative)인 정보들을 활용해 산출되어 왔다.

* 이 논문 또는 저서는 2014년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (NRF-2014S1A5A2A03064791)

이처럼 정량적이고 명시적인 상품에 대한 평점이나 구매 여부와 같은 정보는 수리적으로 처리하기 쉽다는 장점이 있지만, 과연 이러한 정보들이 진정으로 고객의 선호체계를 대표할 수 있는가에 대해서는 의문이 제기되고 있다(Zhang et al., 2014). 예를 들어, 스마트폰 어플리케이션(application, 이하 앱)을 이용하는 '갑'과 '을'이라는 사용자가 '카카오톡' 앱에 대해 모두 5점의 별점을 주었다고 했을 때, 두 사용자는 해당 앱에 대해 동일한 선호도를 보인 것이 된다. 하지만, '갑'은 '카카오톡'의 '사용자 편의성'과 '텍스트 채팅' 기능이 좋아서 5점의 별점을 부여한 반면, '을'은 '카카오톡'의 사용자 편의성은 불편하다고 생각했지만 '음성 보이스톡' 기능이 편하고, '사진 및 동영상 공유'가 유용해서 5점의 별점을 부여한 것이라면, 이 두 사용자가 '카카오톡' 앱에 대해 가지고 있는 선호체계는 동일하다고 보기 어렵다. 때문에, 사용자가 '왜' 해당 상품을 선호하거나 구매했는지는 모르는 상태에서, 결과적으로 얼마나 선호했는지나 구매 여부만을 고려하여 협업 필터링의 고객 간 유사도를 산출하는 것은 정확성을 담보하기 어려우며, 이는 협업 필터링 기반 추천 시스템의 추천 정확도를 떨어뜨리는 하나의 중대한 원인이 될 수 있다.

인터넷 보급이 늘어나고 활성화 되면서 웹을 통한 쇼핑으로 고객들의 구매패턴이 변화하고 있다(Shin et al. 2012). 이런 가운데 최근 Web 2.0의 도래와 함께, 텍스트를 이용해 인터넷 상에 자신의 의견을 표출하는 것이 점차 보편화 되어 가고 있다(Chen et al., 2007). 쇼핑몰과 같은 상거래 플랫폼에서도 상품에 대한 고객들의 의견을 공유할 수 있도록 하는 사용자 리뷰가 크게 활성화되고 있는데, 이와 같은 리뷰에는 해당 상품에 대해 고객이 갖고 있는 선호에 대한 보다 상세하고, 신뢰할 수 있는 정보를 담고 있어 추천 시스템에서 활용하기에 매우 유용할 수 있다(Choeh et al. 2013). 이러한 배경에서 본 연구는 정성적(qualitative)인 사용자가 직접 작성한 리뷰를 참고하여, 전통적인 협업 필터링의 정확도를 개선할 수 있는 새로운 추천 알고리즘을 제안한다.

본 연구에서 새롭게 제안하는 추천 알고리즘은 협업 필터링에서 사용자 간 유사도를 산출할 때, 단순히 상품에 대한 해당 사용자의 평점 패턴의 유사성만 고려하는 것이 아니라, 사용자 간 리뷰의 유사성까지 함께 자동으로 고려할 수 있도록 설계되었다. 이 때 정성적인 사용자 리뷰를 정량적으로 분석하기 위해 텍스트 마이닝 기술을 사용하였다. 텍스트 마이닝을 통해 리뷰에서 사용된 단어(색인어)들이 자동으로 식별되면, 특정 상품에 대한 사용자 리뷰에 공통적으로 사용된 단어의 수가 얼마나 되는지, 혹은 공통적으로 사용된 단어들의 영향력이 얼마나 되는지를 산출하여 이를 기반으로 사용자 리뷰 간 유사도를 산출할 수 있도록 하였다. 본 연구에서는 제안된 추천 알고리즘을 가상으로 구축된 실제 스마트폰 앱 마켓플레이스(app marketplace)의 앱 추천 시스템에 적용하여, 그 성능을 실제적으로 검증해 보고자 하였다.

이후 논문의 구성은 다음과 같다. 우선 2장에서는 협업 필터링과 텍스트 마이닝 관련 기본 개념에 대해서 살펴보고, 현재까지 수행된 관련 연구들에 대해 검토한다. 이어 3장에서는 본 연구에서 제안하는 새로운 추천 알고리즘에 대해서 설명한다. 4장에서는 제안 알고리즘의 성능을 검증하기 위한 데이터에 대한 소개와 함께, 실험 설계 및 실험 결과에 대하여 기술한다. 마지막 5장에서는 전체적인 연구성과와 의의를 요약한 뒤, 본 연구의 한계 및 향후 연구 방향을 제시한다.

2. 이론적 배경

본 장에서는 제안 시스템과 관련이 있는 추천 시스템과 협업 필터링, 그리고 텍스트 마이닝의 기본 개념과 원리를 살펴본다. 이어 사용자 리뷰를 활용한 추천 시스템 관련 연구 동향을 살펴보고, 국내·외 기존 연구들의 시사점과 한계를 살펴보게 될 것이다.

2.1 추천 시스템

추천 시스템(recommender system)이란 특정 사용자를 위한 Top-N 추천 상품 목록을 생성하거나 추천 대상 상품들에 대한 해당 사용자의 평가점수를 예측하는 방법을 통해, 그들이 전자 상거래 사이트에서 구매를 희망하는 상품을 쉽게 찾을 수 있도록 도와주는 데이터 분석 기술 기반의 정보 여과(information filtering) 시스템을 말한다(Sarwar et al., 2001). 대부분의 추천 시스템은 내용 기반(content-based, CB) 또는 협업 필터링(collaborative filtering, CF) 알고리즘에 기반을 두고 있다. 이 중 보편적으로 더 많이 사용되는 추천 알고리즘은 협업 필터링이다. 상품 간 유사성(item-to-item similarity)에 기반하여 추천결과를 생성하는 내용 기반 추천 알고리즘과 달리, 협업 필터링은 사용자 간 유사성(user-to-user similarity)에 기반하여 추천 결과를 생성하는데, 일반적으로 내용 기반 추천 알고리즘에 비해 상대적으로 더 우수한 추천 정확도를 보이는 것으로 알려져 있다(Kim and Ahn, 2009; Kim and Ahn, 2011).

협업 필터링 알고리즘에는 크게 메모리 기반과 모델 기반의 2가지 접근 유형이 있다(Breese et al., 1998). 메모리 기반 협업 필터링은 추천 상품 산출 시, 전체 사용자-상품 점수 행렬(user-item matrix)의 원본을 그대로 사용한다. 이 접근법에서는 추천 대상이 되는 사용자와 유사한 패턴을 가진 것으로 파악된 사용자 그룹에 속한 모든 점수의 가중치가 부여된 평균을 활용하여 추천결과를 생성한다. 메모리 기반 시스템의 주요 장점은 적용이 용이하고, 추천결과에 대한 설명이 가능하다는 점이다. 반면 추천결과를 생성할 때마다 매번 많은 연산이 요구된다는 단점을 가진다.

한편 모델 기반 협업 필터링은 베이지안 네트워크(Bayesian network)나 군집분석(clustering) 등을 통해 사용자-상품 점수 행렬을 기반으로 사용자 등급을 설명하는 모형을 개발(학습)한 다음, 해당 모형을 기반으로 상품을 추천하는 방식을 취한다. 이 접근법의 경우, 모형을 학습하는데에는 시간과 연산량이 많이 소요되지만, 한 번 모형을 학습해 놓고 나면 그 다음 적용을 통해 추천결과를 생성하는데에는 시간과 연산량이 많이 소요되지 않는다는 장점이 있다. 다만, 추천 정확도가 메모리 기반 접근법에 비해 다소 떨어질 수 있고, 사용자들의 선호가 빠르게 또는 자주 갱신되어야 하는 환경에 적합하지 않다는 단점이 있다(Schafer et al., 2001).

일반적으로 많이 사용되는 메모리 기반 협업 필터링 알고리즘은 다음 세 단계 절차를 통해 특정 상품에 대한 추천 대상 사용자의 평가점수를 예측한다.

단계 1. 사용자 간 유사도 계산

우선 1단계에서는 추천 대상이 되는 사용자와 다른 사용자들 사이의 유사도를 산출하는 작업이 수행된다. 이러한 사용자 간 유사도에는 피어슨 상관계수(Pearson correlation coefficient, 이하

PCC) 혹은 코사인 유사도(cosine similarity)가 주로 사용된다. 이 중, PCC를 이용하여 사용자 간 유사도를 산출하는 식은 다음 식 (1)과 같다.

$$S_{(x,y)} = \frac{\sum_i (r_{x,i} - \overline{r_x}) \cdot (r_{y,i} - \overline{r_y})}{\sqrt{\sum_i (r_{x,i} - \overline{r_x})^2} \cdot \sqrt{\sum_i (r_{y,i} - \overline{r_y})^2}} \quad (1)$$

상기 식에서 $S_{(x,y)}$는 사용자 x와 사용자 y의 유사도이고, i는 사용자 x와 사용자 y가 공통으로 평가한 상품의 인덱스(index)이다. $r_{x,i}$은 상품 i에 대한 사용자 x의 평가점수이고, $r_{y,i}$은 상품 i에 대한 사용자 y의 평가점수이다. $\overline{r_x}$은 사용자 x의 평가점수 평균값이고, $\overline{r_y}$은 사용자 y의 평가점수 평균값이다.

단계 2. 이웃 선택

1단계에서 추천 대상이 되는 사용자와 다른 모든 사용자들 간의 유사도가 산출되고 나면, 이 유사도를 기반으로 추천 대상 사용자와 가장 유사한 N명의 이웃을 2단계에서 선택하게 된다.

단계 3. 평가점수 예측

마지막 3단계에서는 앞 단계에서 선택된 이웃들의 평가점수를 기반으로 추천 대상 사용자의 평가점수를 예측하는 작업이 수행된다. 이 때 상품 i에 대한 사용자 x의 평가점수인 $p_{x,i}$는 다음 식 (2)에 의해 산출된다.

$$p_{x,i} = \overline{r_x} + \sum_{z \in N} (r_{z,i} - \overline{r_z}) \cdot \frac{s_{x,z}}{\sum_{z \in N} |s_{x,z}|} \quad (2)$$

위 식에서 $\overline{r_a}$는 사용자 a의 평가점수 평균값이고, $s_{x,z}$는 추천 대상 사용자 x와 이웃 사용자 z사이의 유사도를 나타낸다. 그리고 N은 2단계에서 선택된 가장 가까운 이웃들의 집합을, z는 각각의 이웃을 나타내는 인덱스를 의미한다.

2.2 텍스트 마이닝

텍스트는 가장 보편화된 정보를 표현하고 전달하는 수단으로, 인류의 역사가 진행되는 동안 축적되어 온 방대한 양의 지식들이 지금껏 텍스트 형태로 저장되어 왔다(Witten, 2004). 텍스트 마이닝은 이처럼 우리가 흔하게 접할 수 있는 자연어로 구성된 대량의 비정형 텍스트 데이터에서 숨겨진 패턴 또는 관계를 추출하여 의미 있고 활용가치가 높은 정보 또는 지식을 추출하는 일련의 분석 기법을 의미한다(Hearst, 1999; Hyun et al. 2013; Sebastiani, 2002).

기술적 관점에서 텍스트 마이닝은 자연어 처리 기술을 기반으로 한다. 자연어 처리 기술은 언어 현상을 기계적으로 분석해서 컴퓨터가 이해할 수 있는 형태로 만들거나, 그러한 형태를 다시 인간이 이해할 수 있는 자연언어로 표현하는 기술을 의미한다. 자연어 처리의 대상이 되는 텍스트는 목적에 따라 다양한 형태로 표현될 수 있지만, 일반적으로 벡터공간모델(Vector Space Model)을 이용하여 표현한다(Salton et al., 1975). 벡터공간모델을 이용하게 되면, 문서별 사용된 용어의 빈도수에 따라 문서의 주제 및 특정 단어가 손쉽게 요약된다.

텍스트 마이닝에서는 용어를 분석할 때 단순 빈도수보다는 TF-IDF(Term Frequency-Inverse document Frequency)에 근거한 분석을 주로 활용한다. TF-IDF는 정보 검색과 텍스트 마이닝에서 이용하는 가중치로, 여러 문서로 이루어진 문서

군이 있을 때 어떤 단어가 특정 문서 내에서 얼마나 중요한 것인지를 나타내는 통계적 수치를 의미한다. TF-IDF 가중치는 문서의 핵심어를 추출하거나, 검색 엔진에서 검색 결과의 순위를 결정할 때, 혹은 문서 간 유사도를 산출하는 등의 용도로 사용될 수 있다(Salton and McGill, 1986; You et al. 2004).

TF(term frequency)는 특정 단어가 문서 내에 얼마나 많이 출현하는지를 나타내는 빈도값이다. 일반적으로 TF가 높으면, 해당 단어는 문서 내에서 중요한 단어일 확률이 높다. 하지만, 해당 단어가 만약 일반적으로 자주 쓰이는 단어라면, 어떤 문서 내에서 자주 출현했다고 해서 그 단어가 해당 문서에서 중요한 단어라 판단하기 어려워진다. 이러한 문제를 보정하기 위해 도입된 개념이 IDF(inverse document frequency)이다. IDF는 DF(document frequency)의 역수를 의미하는데, 전체 문서 중에 특정 단어를 포함한 문서의 비중이 높을수록 반비례하여 그 값이 낮아지게 된다. 실질적으로 IDF는 전체 문서의 수를 해당 단어를 포함한 문서의 수로 나눈 값에 로그(log)를 취하여 산출한다. 그리하여, 최종적인 특정 단어에 대한 TF-IDF 가중치는 TF와 IDF를 곱한 값으로 산출된다.

2.3 사용자 리뷰를 반영한 추천 시스템

추천 시스템에 사용자 리뷰를 반영하고자 하는 시도들이 지난 2000년대 중반 이후부터 학계에 소개되고 있다. 그 중 가장 최초의 시도라 볼 수 있는 연구는 Leung et al.(2006)이다. 이 연구에서 저자들은 IMDb(www.imdb.com) 영화 리뷰에 감성 분석(sentiment analysis)을 적용하여, 사용자 리뷰가 긍정적인지 혹은 부정적인지에 대한 감성 방향(sentimental orientation)과 해당 의견의 강도(intensity)를 추정할 수 있는 확률적 평가 추론 모형(probabilistic rating inference model)을 새로 개발하고, 모형의 산출 값을 기반으로 작동하는 협업 필터링 기반의 추천 시스템을 제안하였다. 이 연구는 사용자 리뷰라는 정성적 정보를 정량적으로 해석하여 추천 시스템에 접목을 시도한 첫 번째 연구라는 점에서 그 의의가 있다. 하지만, 제안된 추천 시스템이 사용자가 납득할 만한 추천결과를 실제로 만들어 낼 수 있는지에 대해서는 정확하게 검증되지 못했다는 점과 정량적 평가점수와 정성적인 사용자 리뷰를 동시에 고려하면 더 성능 좋은 추천 알고리즘을 제안할 수 있음에도 불구하고, 정량적 평가점수를 전혀 고려하지 않았다 한계를 갖는다.

한편 Jacob et al.(2009)은 사용자 리뷰를 활용한 모델 기반 하이브리드 협업 필터링 시스템을 제안하였다. 이들은 사용자 리뷰로부터 추출된 다양한 상품 특성에 관한 사용자 의견을 기반으로 다중 관계형 모델(multi-relational model) 형식의 사용자 프로파일을 구축한 다음, 이를 기반으로 협업 필터링을 적용하여 추천결과를 생성하게끔 하였다. 제안 기법의 검증을 위해 IMDb 데이터를 수집해 적용하였으며, 실험결과 평가점수 기반의 전통적인 협업 필터링보다 95% 신뢰수준 하에서 통계적으로 유의하게 성능이 개선됨을 확인할 수 있었다.

Wang et al.(2012)은 Jacob et al.(2009)이 제안한 추천 알고리즘의 기본 틀을 그대로 채택하되, 텐서 분해(tensor factorization)라는 새로운 기법을 통해 보다 효과적으로 사용자의 평가점수를 예측할 수 있는 방법을 제안하였다. 그리고 실험을 통해 기존에 소개된 기법들에 비해 이들의 기법이 보다 정확한 예측결과를 산출함을 검증하

였다. Jacob et al.(2009)과 Wang et al.(2012)의 연구는 정량적 평가점수 기반의 추천 알고리즘에 사용자 리뷰 마이닝 결과를 추가로 반영하여 성능이 더 개선됨을 보였다는 점에서 의의가 있다. 하지만, 이들의 접근법은 극성을 파악해야 하는 오피니언 마이닝에 기반하고 있어 리뷰에 대한 전처리 과정이 복잡하고, 그 과정에 수작업이 일부 요구된다는 한계를 갖는다.

Moshfeghi et al.(2011)과 Levi et al.(2012)는 협업 필터링의 주요한 문제 중 하나인 신규 고객 추천 문제(cold start problem)을 해결하기 위한 대안으로 사용자 리뷰 마이닝을 활용하는 방안을 제안하였다. 이들은 제안 모형을 각각 영화와 호텔 추천에 적용하고, 실증 분석을 통해 제안 모형이 신규 고객에게 만족할만한 추천결과를 생성함을 확인하였다.

한편 Garcia-Cumbreras et al.(2013)은 사용자 리뷰에 감성 분석을 적용하여, 해당 사용자가 일반적으로 긍정적인 평가를 주로 하는 낙관주의자(optimist)인지 혹은 부정적인 평가를 주로 하는 비관주의자(pessimist)인지를 분류하고, 각 집단별로 따로 협업 필터링을 수행하도록 하는 새로운 접근의 추천 시스템을 제안하였다. 저자들은 IMDb를 이용한 실증분석을 통해, 전체 사용자를 대상으로 협업 필터링을 수행할 때 보다 낙관주의자와 비관주의자로 나누어 협업 필터링을 수행할 때 더 낮은 예측오차를 보임을 확인하였다. 이 연구는 사용자 리뷰를 집단 분류의 기준으로 사용함으로써 이전의 연구들과 차별화된 접근을 시도하고 있다는 점에서 의의가 있지만, 리뷰의 내용을 직접적으로 추천 알고리즘에 반영하지 못해 정보 손실(information loss)이 발생한다는 한계를 갖는다.

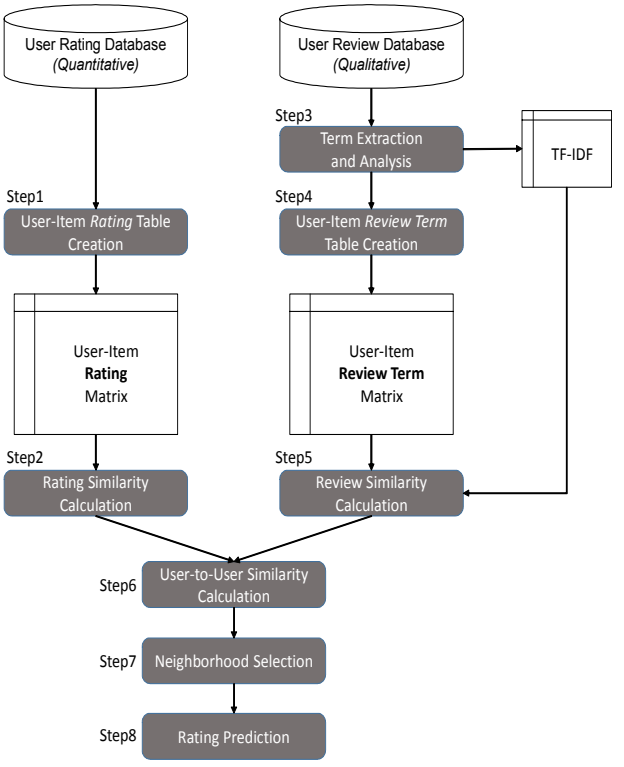
가장 최근에 발표된 Zhang et al.(2014)은 urCF(User Review enhanced Collaborative Filtering)이라는 이름의 메모리 기반 추천 시스템을 제안하였다. 이들은 Zhou and Chaovalit (2008)의 영화 리뷰 온톨로지에 기반하여 총 32개의 영화 리뷰 특징을 도출한 뒤, 각 특징별 사용자 의견의 극성(opinion orientation)을 추가로 반영하여 보다 정밀하게 사용자 간 유사도를 산출하고자 하였다. 이 때, 각 특징별 가중치는 TF-IDF와 사실상 유사한 개념인 FF-IRF(Feature Frequency-Inverse Review Frequency) 가중치를 사용할 것을 제안하였다. Yahoo! Movies로부터 추출된 데이터를 이용해 제안 시스템의 성능을 검증한 저자들은 제안 시스템이 전통적인 협업 필터링 시스템에 비해 6.18～8.24% 가량 예측 정확도를 개선시킴을 확인하였다. 이 연구의 경우, Jacob et al.(2009)이나 Wang et al.(2012)과 달리 메모리 기반 추천 시스템을 제안하여 구현이 용이하고, 예측 정확도가 상대적으로 우수하다는 장점을 갖고 있다. 하지만, 상기 두 연구와 마찬가지로 오피니언 마이닝에 기반하여 리뷰에 대한 전처리 과정이 복잡하고 잘못된 극성 파악이 이루어질 위험이 높다는 점, 특징 도출이 수작업으로 이루어져 영화가 아닌 다른 영역에 쉽게 적용하기 어렵다는 점 등의 한계가 있다. 아울러, 지금까지 살펴본 바와 같이 사용자 리뷰를 활용한 추천 시스템 관련 해외 연구들의 대다수가 IMDb나 Yahoo! Movies와 같은 영화를 대상으로 검증을 수행하고 있어, 다른 분야로의 적용가능성이 충분히 검증되지 못했다는 점은 공통된 한계로 지적될 수 있다.

이처럼 사용자 리뷰를 추천 시스템에 응용하려는 연구들이 해외에서 최근 활성화되고 있는 추세지만, 안타깝게도 국내에서는 거의 찾아보기 힘들다. 사용자 리뷰의 신뢰도를 정량적으로

측정할 수 있는 모형을 제안하고, 이를 추천 시스템에 반영하여 추천 정확도의 개선을 실증적으로 확인한 Choeh et al.(2013)의 연구나 개봉 첫 주 동안의 영화 사용자 리뷰들을 기반으로 영화 흥행 성적 예측모형을 제안한 Cho et al.(2014)의 연구들이 사용자 리뷰의 응용을 시도하였으나, 사용자 리뷰의 내용을 분석하여 이를 추천 시스템에 활용한 사례는 아직 발표되지 않은 상태이다.

3. 제안 알고리즘

본 연구는 텍스트 마이닝을 활용한 사용자 리뷰 유사도를 추가로 반영하여, 메모리 기반 협업 필터링의 성능을 개선할 수 있는 새로운 하이브리드 추천 알고리즘을 제안한다. 본 연구의 제안 알고리즘은 <Figure 1>과 같이 총 8단계에 의해 구동된다.



〈Figure 1〉 Procedure of the proposed algorithm

1단계. 사용자-상품 평가점수 행렬 도출

추천 알고리즘의 첫 단계는 고객 데이터베이스와 거래 데이터베이스에 저장된 사용자들의 상품에 대한 평점 데이터들을 추출·정리하여, 다음의 <Figure 2>에 예시된 것과 같은 사용자-상품 평점 행렬(User-Item Rating Matrix)을 도출하는 것이다. <Figure 2>를 통해 볼 수 있듯이, 사용자-상품 평점 행렬은 각 사용자들의 상품 평점 패턴이 행 단위로 정렬되어 있어, 사용자 간 유사도를 산출할 때 편리하게 참조될 수 있다.

	Item 1	Item 2	Item 3	...	Item *m*
User 1		5	4		1
User 2	4		3		4
User 3	3	2			3
...					
User *n*-1	2		3		1
User *n*	5	1	3		4

Ratings

〈Figure 2〉 Example of user-item rating matrix

2단계. 사용자 간 평가점수 유사도 산출

2단계에서는 앞서 1단계를 통해 도출된 사용자-상품 평가점수 행렬을 기반으로 추천 대상 사용자와 다른 사용자들 간의 평가점수 유사도를 산출하게 된다. 평가점수 유사도는 일반적으로

피어슨 상관계수(PCC) 혹은 코사인 유사도를 이용해 산출되는데, 본 연구의 제안 알고리즘은 PCC에 기반한 평가점수 유사도를 사용한다. PCC를 활용한 두 사용자 간 평가점수 유사도는 앞서 2장 1절에서 소개한 식 (1)에 의해 산출된다.

3단계. 사용자 리뷰 색인어 추출 및 분석

1~2단계의 작업을 통해 정량적 정보인 평가점수를 활용한 사용자 간 유사도 산출이 완료되면, 3~5단계에서는 정성적 정보인 사용자 리뷰를 활용해 다시 한 번 사용자 간 유사도를 산출하게 된다. 그 첫 번째 활동인 3단계에서는 고객 데이터베이스와 거래 데이터베이스에 저장된 각 상품에 대한 사용자들의 리뷰를 추출한 뒤, 자연어 처리 기술인 색인어 추출(term extraction)을 적용해 이를 색인어들의 집합으로 변환하는 작업을 수행하게 된다.

그런 다음 추출된 모든 색인어들에 대해 TF-IDF 분석을 적용하여, 각 색인어의 리뷰 내 영향력을 산출한다. TF-IDF 가중치의 산출 원리는 본 논문 2장 2절에 설명되어 있다.

4단계. 사용자-상품 리뷰 단어 행렬 도출

앞선 단계를 통해 리뷰를 색인어들의 집합으로 변환하고 난 다음에는 이를 <Figure 3>에 제시된 사용자-상품 리뷰 단어 행렬(User-Item Review Term Matrix)의 형태로 재구성하게 된다. 사용자-상품 리뷰 단어 행렬은 1단계의 사용자-상품 평점 행렬과 마찬가지로 다음 단계에서 사용자 간 리뷰 유사도를 산출하는데 활용된다.

	Item 1	Item 2	...	Item *m*
User 1		{편의, 훌륭, 속도...}		{별로, 없다, 불만...}
User 2	{디자인, 우수, ...}			{예술, 환상, 느리다...}
User 3	{가격, 성능, 추천...}	{친구, 성능, 업뎃...}		{무난, 색상, 위치...}
...				
User *n*-1	{실망, 성능, 속도...}			{회사, 반성, 어제...}
User *n*	{관심, 활용, 가족...}	{돈, 아깝다, 시간...}		{지인, 추천, 대박...}

Review Words (Vector)

〈Figure 3〉 Example of user-item review term matrix

5단계. 사용자 리뷰 유사도 산출

이 단계에서는 앞 단계에서 도출한 사용자-상품 리뷰 단어 행렬을 기반으로, 사용자 간 리뷰 유사도를 산출하게 된다. 그런데, 사용자-상품 리뷰 단어 행렬의 경우, 사용자-상품 평점 행렬과 달리 행렬 내 원소의 값이 단어 벡터로 구성되어 있어, PCC나 코사인 유사도 같은 방식으로 두 사용자 간 유사도를 산출하는 것이 불가능하다.

이에 본 연구에서는 2가지 새로운 방법을 그 대안으로 제시한다. 우선 첫 번째 방법은 동일한 아이템에 대한 두 사용자 간 리뷰에 얼마나 많은 단어가 공통적으로 포함되었는지 그 빈도를 세어 이 값을 유사도로 활용하는 것이다. 예를 들어, *i*번째 상품에 대한 A 사용자의 리뷰에 {가격, 성능, 만족, 최고, 구입}이라는 색인어가 추출되었고, 동일 상품에 대한 B 사용자의 리뷰에 {가격, 편의성, 만족, 우수, 추천} 이라는 색인어가 추출되었다면, *i*번째 상품에 대한 A 사용자와 B 사용자의 리뷰 유사도는 n({가격, 만족}) = 2가 된다.

사용자 리뷰 유사도를 도출하는 두 번째 방법은 방금 소개한 첫 번째 방법을 한 단계 더 확장한 방법으로, 두 사용자가 공통으로 사용한 단어들의 TF-IDF의 가중치 합을 유사도로 사용하는 방법이다. 앞서 소개한 예에서 만약 '가격'이라

는 단어의 평균 TF-IDF 가중치가 3이고, '만족'이라는 단어의 평균 TF-IDF 가중치가 5라면, TF-IDF를 고려한 A와 B 두 사용자의 *i*번째 상품에 대한 리뷰 유사도는 3 + 5 = 8이 된다.

전자의 경우 적용이 간편하다는 장점이 있고, 후자의 경우 좀 더 정밀하게 리뷰 유사도를 산출할 수 있다는 장점이 있다. 이에 본 연구에서는 이 두 방법을 모두 사용해 보고, 그 성능을 비교해 보고자 한다. 또한 본 연구에서 제시한 첫 번째 방법과 두 번째 방법 모두 상황에 따른 편차가 크게 나타날 수 있어, 사용 시 적절한 보정이 요구된다. 이에 본 연구에서는 산출된 유사도 값을 최대값으로 나누어, 어떤 방법을 사용하더라도 특정 상품에 대한 두 사용자 간 리뷰 유사도 값이 항상 0~1 사이의 값을 갖도록 보정하는 방법을 사용하였다.

6단계. 사용자 간 유사도 산출

2단계와 5단계를 통해 사용자 간 평점 유사도와 사용자 간 리뷰 유사도가 산출되고 나면, 6단계에서는 이 두 유사도를 통합하여 전체적인 사용자 간 유사도를 산출하게 된다. 이 때, 사용자 간 유사도 $S_{x,y}$는 다음 식 (3)과 같이 산출된다.

$$S_{x,y} = s_{x,y}^{Rate} \times (1 + \rho \cdot s_{x,y}^{Review}) \quad (3)$$

위 식에서 $s_{x,y}^{Rate}$는 사용자 *x*, *y*의 평점 유사도(-1~1), $s_{x,y}^{Review}$는 사용자 *x*, *y*의 리뷰 유사도(0~1), 그리고 ρ는 조정계수를 의미한다. 여기서 조정계수는 리뷰 유사도를 전체 유사도에 어느 정도 비중 있게 반영할 것인가를 결정하는 지표로서, 모형 설계자가 시행착오(trial-and-error)를 통해 최적값을 찾아야 하는 모수이다.

7단계. 이웃 선정

6단계를 통해 추천 대상 사용자와 다른 사용자 간의 유사도가 모두 산출되고 나면, 이 중에서 가장 가까운 이웃 집단 *N*을 선정하게 된다.

8단계. 예상 평가점수 산출 및 추천 대상 상품 선정

마지막 8단계에서는 7단계에서 선정된 이웃 집단 *N*을 기반으로 추천 대상 사용자의 상품별 예상 평가점수(만족도)를 산출하는 작업이 이루어진다. 이 때 예상 평가점수 산출은 앞서 2장 1절에 소개된 식 (2)에 의해 산출된다. 이 작업이 끝나게 되면 추천 시스템은 최종적으로 예상 평가점수가 높은 상품들 중에서, 추천 대상 사용자가 아직 경험해 보지 못한 상품들을 중심으로 추천 대상 상품을 선정하게 된다.

4. 실험설계 및 실증분석

4.1 데이터 수집

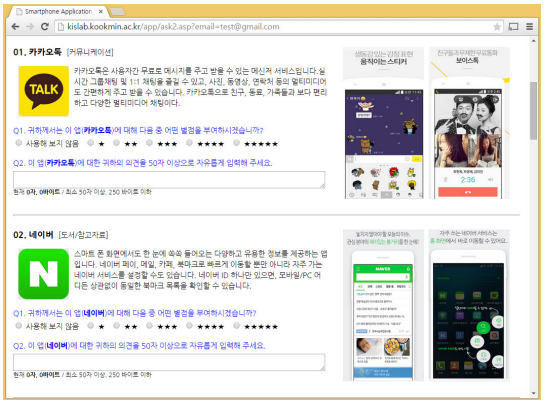
제안한 추천 알고리즘의 성능을 검증하기 위하여, 본 연구에서는 이를 국내 스마트폰 사용자의 앱 추천시스템 사례에 적용해 보았다. 최근 스마트폰의 급격한 보급으로 인해, 앱 마켓플레이스에 등록되어 거래되는 앱의 수도 엄청나게 증가하고 있다. 2014년 7월 기준으로 애플 앱 스토어에는 120만종, 구글 플레이에는 무려 130만종 가량의 앱이 제공(Statista, 2015)되고 있는데, 이처럼 과도하게 많은 앱이 현재 제공되고 있어 사용자에게 적절한 앱을 추천해 줄 수 있는 추천시스템의 개발은 실무적으로 대단히 중요한 의미가 있다. 특히, 구글 플레이나 애플 앱 스토어

에는 <Figure 4> 또는 <Figure 5>와 같이 앱에 대한 사용자 평가 기능을 오래 전부터 제공되고 있어, 관련 데이터가 충분히 축적되어 있는 상태이다. 때문에, 우리는 스마트폰 앱이 본 연구에서 제안하는 사용자 리뷰를 반영한 추천 알고리즘의 성능을 점검하기에 가장 최적화된 연구 대상이라 판단하고, 이를 대상으로 연구를 진행하였다.

〈Figure 4〉 User’ s ratings in practice(Google)

〈Figure 5〉 User’ s ratings in practice(Apple)

하지만, 구글이나 애플과 같은 글로벌 기업의 협조를 통해 데이터를 확보하는 것이 현실적으로 한계가 있어, 실험 데이터를 확보하기 위한 다른 방법이 요구되었다. 이에 본 연구팀에서는 <Figure 6>과 같은 온라인 설문 시스템을 구축한 뒤, 구글 플레이 및 애플 앱 스토어에 모두 등록되어 있는 앱 중 지난 2014년 한 해 동안 국내에서 가장 인기를 끌었던 총 48종 앱에 대해 설문 참여자들의 평점(별 5개 만점)과 최소 50자 이상의 의견을 수집하였다. 설문 시 각 앱에 대한 간략한 설명과 스크린샷을 함께 제공하여, 설문 응답자가 해당 앱을 쉽게 인지할 수 있도록 지원하였으며, 설문 참여자가 중간에 응답을 중단해도 언제든 남은 설문들을 나중에 완성할 수 있도록 설계하여 최대한 충실한 설문 응답이 이루어질 수 있도록 최선의 노력을 기울였다.



〈Figure 6〉 Sample screen of the Web-based data collection

2015년 4월 15일부터 이틀간 10명을 대상으로 선행 테스트를 먼저 수행한 뒤, 본 설문을 4월 17일부터 4월 30일까지 약 2주간 진행하였다. 해당 기간 중 총 98명이 참여했는데, 그 중 18명은 설문을 끝까지 완료하지 않았고, 2명은 불성실하게 응답한 것으로 확인되었다. 이에 20명의 응답을 제외한 총 78명의 설문 데이터를 기반으로 제안 알고리즘에 대한 검증을 수행하였다.

분석 대상자의 평균 연령은 24.4세로 대학생 및 대학원생 응답자가 대다수를 이루었으며, 성별의 경우 남성이 44명, 여성이 34명으로 남성 응답자의 비중이 약간 더 높았다. 사용하는 스마트폰의 경우, 아이폰 계열이 22명, 안드로이드

계열이 56명으로 나타나 안드로이드폰 사용자의 비중이 2배 이상 높게 나타났으며, 전체 응답자 중 69명이 최소 15개 이상의 스마트폰 앱을 설치해 사용하고 있는 것으로 나타났다. 이 중 28명은 30개 이상의 앱을 설치하고 있다고 응답하여, 상당히 많은 수의 이용자들이 다수의 앱을 설치하여 사용하고 있음을 알 수 있었다.

78명 응답자의 전체 응답건수는 1,246건으로, 응답자 1인당 평균 15.97개의 앱에 대해 평점과 리뷰를 응답한 것으로 조사되었다. 결국, 사용자-상품 행렬의 전체 셀 중 약 33.28%만 채워지게 되어, 상당히 희박성(sparsity)이 높은 상황에서 제안된 추천 알고리즘의 성능을 검증하게 되었다고 할 수 있다.

4.2 실험설계

본 연구에서는 제안한 추천 알고리즘을 Microsoft Excel에 내장된 VBA(Visual Basic for Applications) 프로그래밍 언어로 직접 구현하여 실험을 수행하였다. 단, 사용자 리뷰로부터 색인어를 추출하고 TF-IDF 가중치를 산출하는 작업은 SAS Text Miner를 활용하였다. 텍스트 마이닝 수행 시, 추출된 색인어들을 사람이 개입하여 정제할 경우 성능이 왜곡될 위험이 있음을 우려하여, SAS Text Miner가 자동으로 추출한 색인어들을 별도의 후처리 작업 없이 그대로 사용하였다.

제안 알고리즘을 구현하기 위해서는 사용자 리뷰 유사도를 계산할 때, 사용자 간 공통으로 출현한 단어들의 빈도나 TF-IDF 가중치의 합을 최대값을 이용하여 정규화하는 작업이 요구된다. 따라서, 빈도와 TF-IDF 가중치의 합에 대한 최대값을 찾아야 하는데, 본 연구에서 사용한 데이터에서 실제 최대값은 빈도의 경우 44, TF-IDF 가중치 합의 경우 59.18로 나타났다. 하지만, 이 값들을 그대로 사용할 경우 극단치(outlier)에 전체적인 비율값이 영향을 받게 될 것으로 우려되었다. 이에 실험 수행 시, 빈도의 경우 5～10, TF-IDF 가중치는 10～25 사이의 값들을 최대값으로 설정해 보고 가장 예측성과가 우수한 값을 탐색하는 방식을 취하였다. 만약 지정한 최대값보다 큰 값의 빈도가 TF-IDF 가중치 합을 갖게 되는 경우에는 강제로 최대값을 할당하는 윈저라이징(winsorizing) 방법을 적용하였다. 아울러, 조정계수 ρ 역시 1～20 사이의 값을 적용해 보면서, 어떤 값을 가질 때 가장 성과가 우수한 지 탐색하고자 하였다.

한편 제안 알고리즘이 과연 얼마나 성능의 개선을 가져오는지 확인해 보기 위해, 전통적인 협업 필터링 알고리즘을 비교 알고리즘으로 정하여 함께 실험하였다. 아울러, 제안 알고리즘의 경우, (1) 공통으로 출현한 색인어의 빈도로 사용자 리뷰 유사도를 산출하는 방식과 (2) 공통으로 출현한 색인어의 TF-IDF 가중치 합으로 사용자 리뷰 유사도를 산출하는 방식으로 모두 적용해 보고, 그 성능을 비교해 보고자 하였다.

4.3 실험결과

본 연구에서는 All-but-One 방식으로 실제 평점이 입력된 상품에 대해 추천 알고리즘으로 예상 평점을 도출해 본 다음, 실제 평점과의 평균 오차가 가장 적게 나타난 추천 알고리즘이 어떤 것인지를 확인하는 방식으로 검증 작업을 진행하였다(Kim and Kim, 2014). 이 때 실제 평점과 예상 평점 간의 오차는 추천 시스템 관련 연구에서 전통적으로 많이 사용되어 온 평균

MAE(Mean Absolute Error)를 통해 산출하였다 (Breese et al., 1998; Sarwar et al., 2001).

다음의 <Table 1>에 제안 알고리즘과 비교 알고리즘의 성능, 즉 평균 MAE가 제시되어 있다. 이 표에서 볼 수 있듯이 사용자 리뷰를 추가로 고려하는 본 연구의 제안 알고리즘이 평점만을 고려하는 전통적인 협업 필터링에 비해 크진 않지만 어느 정도 성능의 개선을 가져옴을 확인할 수 있었다. 아울러, 제안 알고리즘의 경우 사용자 리뷰 유사도를 공통으로 사용된 단어의 빈도로 산출하는 경우보다 TF-IDF 가중치를 사용해 산출했을 때 예측 정확도가 약간 더 높아짐을 확인할 수 있었다.

그리고 제안 알고리즘에서 최적의 조정계수 ρ가 각각 10, 20으로 상당히 큰 값이 도출되었다는 점도 흥미롭다. 이는 본 연구의 대상이 된 앱 추천에서 평점 유사도보다 리뷰 유사도의 비중을 10배, 20배 더 가중하여 고려했을 때 성능이 더 좋아졌음을 의미하는데, 이를 통해서도 사용자 리뷰 유사도가 무시해서는 안 될 대단히 의미있는 정보원임을 확인할 수 있다.

〈Table 1〉 Experimental results

Recommendation algorithms		Mean of MAE	S.D. of MAE	Optimal Setting
Conventional CF		0.7939	0.3047	-
Proposed algorithm	Freq	0.7881	0.3036	Max: 10, ρ=10
	TF-IDF	0.7867	0.2963	Max: 25, ρ=20

제안 알고리즘이 비교 알고리즘(전통적 협업 필터링)에 비해, 얼마나 유의한 성과의 차이를 보였는지 검증하기 위해 대응표본 t-검정을 수행해 보았다. 그 결과, 빈도를 사용한 제안 알고리즘은 90% 신뢰수준 하에서 MAE의 차이가 서로 유의한 것(t=1.652, p-value=0.10)으로 나타났으나, TF-IDF를 사용한 제안 알고리즘은 그 차이가 통계적 유의성을 확보하지 못하는 것(t=1.325, p-value=0.19)으로 나타났다.

5. 결 론

본 연구에서는 정량적인 평점만을 활용해 추천 결과를 생성하는 기존의 협업 필터링을 개선하기 위해, 정성적인 정보인 사용자 리뷰의 유사도를 추가로 고려하여 협업 필터링의 성능을 높일 수 있는 새로운 추천 알고리즘을 제안하였다. 구체적으로 본 연구에서는 두 사용자의 리뷰에서 공통적으로 추출된 색인어들의 빈도 혹은 해당 색인어들의 TF-IDF 가중치 합으로 사용자 리뷰 유사도를 산출하는 새로운 접근법을 제시하고, 이 두 가지 접근법 중 어느 것이 더 우수한지 실증분석을 통해 확인하였다. 제안 알고리즘을 검증하기 위해, 본 연구에서는 별도의 설문 시스템을 통해 확보한 스마트폰 앱 추천 시스템 데이터에 제안 알고리즘과 전통적인 협업 필터링 알고리즘을 동시에 적용해 보고, 그 성능을 비교해 보았다. 그 결과, 제안 알고리즘이 예측 정확도를 개선시킴을 확인할 수 있었으며, 사용자 리뷰 유사도 산출 시 TF-IDF 가중치를 고려하는 것이 더 나은 추천결과를 생성하는데 기여할 수 있음을 확인하였다.

본 연구는 다음과 측면에서 다양한 학술적·실무적 의의를 갖는다. 우선 본 연구는 국내는 물론, 국외에서도 그간 충분히 다루어지지 않았던

사용자 리뷰 마이닝을 효과적으로 협업 필터링에 접목할 수 있는 방안을 연구했다는 점에서 의미가 있다. 특히 본 연구의 제안 알고리즘은 복잡하고 사람의 인위적 개입이 상대적으로 많이 요구되는 오피니언 마이닝이 아닌 가장 기초적인 수준의 자동화된 자연어 처리 기술인 색인어 추출 기술만으로 구동이 가능한 알고리즘을 제안했다는 점에서 실무적으로 의미가 있는 연구라 사료된다.

또한 데이터 수집이 상대적으로 용이하다는 이유로 대다수의 추천시스템 연구에서 사용되어 온 영화 관련 데이터를 사용하지 않고, 오늘날 그 중요성이 더 크게 대두되고 있는 스마트폰 앱 마켓플레이스 데이터를 따로 수집해, 검증에 사용하였다는 점도 기존 연구와 본 연구가 크게 차별화되는 부분이라고 할 수 있다.

하지만 명확하게 드러난 본 연구의 한계점 역시 지적하지 않을 수 없다. 우선 정성적인 사용자 리뷰를 함께 고려한 추천 알고리즘의 예측 정확도가 비교 알고리즘에 비해 통계적으로 유의한 수준의 개선을 이루어 내지 못했다는 점을 지적할 수 있다. 이는 추천 알고리즘의 예측 정확도 개선이 기대만큼 충분히 이루어지지 못한 것에 기인할 수도 있고, 분석에 사용된 데이터 셋의 규모가 78건으로 너무 작았던 것에 기인할 수도 있다. 때문에 추후 연구에서는 우선 보다 방대한 규모의 실험 데이터로 제안 알고리즘의 성능을 검증해 볼 필요가 있을 것으로 보인다.

아울러, 제안된 추천 알고리즘의 성능을 보다 개선할 수 있는 방안에 대해서도 추가적인 고민이 요구된다. 예를 들어, 사용자 리뷰에 감성 분석을 적용하여 리뷰의 긍·부정 정도를 파악하고, 이러한 리뷰 내용의 전반적인 분위기가 사용자 간에 얼마나 유사한지를 추가로 반영하는 방안을 고려해 볼 수 있다. 이렇듯 주어진 정성적 정보, 즉 사용자 리뷰로부터 보다 많은 유익한 시사점을 추출하여 이를 추천 알고리즘 성능 개선에 반영하려는 노력이 향후 연구에서 더욱 경주되어야 할 것으로 생각된다.

참고문헌(References)

Breese, J. S., D. Heckerman, and C. Kadie, “Empirical analysis of predictive algorithms for collaborative filtering,” *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence,* (1998), 43～52.

Chen, P.-Y., S. Dhanasobhon, and M. D. Smith, “An Analysis of the Differential Impact of Reviews and Reviewers at Amazon.Com,” *Proceedings of International Conference on Information Systems(ICIS)*, (2007), 94.

Choeh, J. Y., S. K. Lee, and Y. B. Cho, “Applying Rating Score’s Reliability of Customers to Enhance Prediction Accuracy in Recommender System,” *Journal of Digital Contents Society*, Vol. 13, No. 7(2013), 379～385.

Cho, S. Y., H.-k. Kim, B. S. Kim, and H.-w. Kim, “Predicting Movie Revenue by Online Review Mining Using the Opening Week Online Review,” *Information Systems Review*, Vol. 16, No. 1(2014), 113～134.

Garcia-Cumbreras, M. A., A. Montejo-Raez, and M. C. Diaz-Galiano, “Pessimists and optimists: Improving collaborative filtering through sentiment analysis,” *Expert Systems with Applications*, Vol. 40, No. 17(2013), 6758～6765.

Hearst, M. A., “Untangling text data mining,"

Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics on Computational Linguistics, (1999), 3～10.

Hyun, Y., H. Han, H. Choi, J. Park, K., Lee, K. -Y. Kwahk and N. Kim, "Methodology Using Text Analysis for Packaging R&D Information services on Pending National Issues," *Journal of Information Technology Applications & Management*, Vol. 20, No. 3(2013), 231～257.

Jacob, N., S. H. Weber, M. C. Muller, and I. Gurevych, "Beyond the Stars: Exploiting Free-Text User Reviews to Improve the Accuracy of Movie Recommendations," *Proceedings of the 1st International CIKM Workshop on Topic-sentiment Analysis for Mass Opinion,* Hong Kong, China, (2009), 57～64.

Kim, K.-j. and H. Ahn, "User-Item Matrix Reduction Technique for Personalized Recommender Systems," *Journal of Information Technology Applications & Management*, Vol. 16, No. 1(2009), 97～113.

Kim, K.-j. and H. Ahn, "Collaborative Filtering with a User-Item Matrix Reduction Technique for Recommender Systems," *International Journal of Electronic Commerce*, Vol. 16, No. 1(2011), 107～128.

Kim, M. and K.-j. Kim, "Recommender Systems using Structural Hole and Collaborative Filtering," *Journal of Intelligence and Information System*, Vol. 20, No. 4(2014), 107～120.

Leung, C. W.-k., S. C.-f. Chan, and F.-l. Chung, "Integrating Collaborative Filtering and Sentiment Analysis: A Rating Inference Approach," *Proceedings of the ECAI 2006 Workshop on Recommender Systems*, Riva del Garda, Italy, (2006), 62～66.

Levi, A., Mokryn, O., Diot, C. and N. Taft, "Finding a Needle in a Haystack of Reviews: Cold Start Context-Based Hotel Recommender System," *Proceedings of the Sixth ACM Conference on Recommender Systems*, Dublin, Ireland, (2012), 115～122.

Moshfeghi, Y., B. Piwowarski, and J. M. Jose, "Handling Data Sparsity in Collaborative Filtering Using Emotion and Semantic Based Features," *Proceedings of the 34th International ACM SIGIR Conference on Research and Development in Information Retrieval*, Beijing, China, (2011), 625～634.

Salton, G. and M. J. McGill, *Introduction to Modern Information Retrieval,* McGraw -Hill, 1986.

Salton, G., A. Wong, and C. S. Yang, "A vector space model for automatic indexing," *Communications of the ACM*, Vol. 18, No. 11 (1975), 613～620.

Sarwar, B., Karypis, G., Konstan, J. and Riedl, J., "Item-based collaborative filtering recommendation algorithms," *Proceedings of the 10th International Conference on World Wide Web*, (2001), 285～295.

Schafer, J. B., J. Konstan, and J. Riedl, "Electronic Commerce Recommender Applications," *Journal of Data Mining and Knowledge Discovery*, Vol. 5, Nos. 1-2(2001), 115～152.

Sebastiani, F., "Machine learning in automated text categorization," *ACM Computing Surveys*, Vol. 34, No. 1(2002), 1～47.

Shin, C. H., J. W.Lee, H. N. Yang, and I. Y. Choi, "The Research on Recommender for New Customers Using Collaborative Filtering and Social Network Analysis," *Journal of*

Intelligence and Information Systems, Vol. 18, No.4(2012), 19～42.

Statista, Number of apps available in leading app stores as of July 2014, 2015. Available at http://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/ (Downloaded 11 May, 2015).

Wang, Y., Y. Liu, and X. Yu, "Collaborative Filtering with Aspect-Based Opinion Mining: A Tensor Factorization Approach," *Proceedings of 2012 IEEE 12th International Conference on Data Mining (ICDM)*, Brussels, Belgium, (2012), 1152～1157.

Witten, I. H., *Text Mining: Practical Handbook of Internet Computing*, CRC press, 2004.

You, M., J.-S. Park, and J.-K. Kim, "Folder Recommendation Based on User Knowledge," *Journal of Intelligence and Information System*, Vol. 10, No. 3(2004), 133～146.

Zhang, Z., D. Zhang, and J. Lai, "urCF: User Review Enhanced Collaborative Filtering," *Proceedings of the 20th Americas Conference on Information Systems*, (2014).

Zhou, L. and P. Chaovalit, , "Ontology-Supported Polarity Mining," *Journal of the American Society for Information Science and Technology*, Vol. 59, No. 1(2008), 98～110.

Abstract

A Collaborative Filtering System Combined with Users' Review Mining : Application to the Recommendation of Smartphone Apps

ByeoungKug Jeon* · Hyunchul Ahn**

Collaborative filtering(CF) algorithm has been popularly used for recommender systems in both academic and practical applications. A general CF system compares users based on how similar they are, and creates recommendation results with the items favored by other people with similar tastes. Thus, it is very important for CF to measure the similarities between users because the recommendation quality depends on it. In most cases, users' explicit numeric ratings of items(i.e. quantitative information) have only been used to calculate the similarities between users in CF. However, several studies indicated that qualitative information such as user's reviews on the items may contribute to measure these similarities more accurately. Considering that a lot of people are likely to share their honest opinion on the items they purchased recently due to the advent of the Web 2.0, user's reviews can be regarded as the informative source for identifying user's preference with accuracy.

Under this background, this study proposes a new hybrid recommender system that combines with users' review mining. Our proposed system is based on conventional memory-based CF, but it is designed to use both user's numeric ratings and his/her text reviews on the items when calculating similarities between users. In specific, our system creates not only user-item rating matrix, but also user-item review term matrix. Then, it calculates rating similarity and review similarity from each matrix, and calculates the final user-to-user similarity based on these two similarities(i.e. rating and review similarities). As the methods for calculating review similarity between users, we proposed two alternatives - one is to use the frequency of the commonly used terms, and the other one is to use the sum of the importance weights of the commonly used terms in users' review. In the case of the importance weights of terms, we proposed

* Master's Candidate, Graduate School of Business IT, Kookmin University
** Corresponding Author : Hyunchul Ahn
Associate Professor, Graduate School of Business IT, Kookmin University
77 Jeongneung-ro, Seongbuk-gu, Seoul 136-702, Korea
Tel: +82-2-910-4577, Fax: +82-2-910-5209, E-mail: hcahn@kookmin.ac.kr

the use of average TF-IDF(Term Frequency - Inverse Document Frequency) weights.

To validate the applicability of the proposed system, we applied it to the implementation of a recommender system for smartphone applications (hereafter, app). At present, over a million apps are offered in each app stores operated by Google and Apple. Due to this information overload, users have difficulty in selecting proper apps that they really want. Furthermore, app store operators like Google and Apple have cumulated huge amount of users' reviews on apps until now. Thus, we chose smartphone app stores as the application domain of our system. In order to collect the experimental data set, we built and operated a Web-based data collection system for about two weeks. As a result, we could obtain 1,246 valid responses(ratings and reviews) from 78 users. The experimental system was implemented using Microsoft Visual Basic for Applications(VBA) and SAS Text Miner. And, to avoid distortion due to human intervention, we did not adopt any refining works by human during the user's review mining process. To examine the effectiveness of the proposed system, we compared its performance to the performance of conventional CF system. The performances of recommender systems were evaluated by using average MAE(mean absolute error).

The experimental results showed that our proposed system(MAE = 0.7867 ~ 0.7881) slightly outperformed a conventional CF system(MAE = 0.7939). Also, they showed that the calculation of review similarity between users based on the TF-IDF weights(MAE = 0.7867) leaded to better recommendation accuracy than the calculation based on the frequency of the commonly used terms in reviews(MAE = 0.7881). The results from paired samples t-test presented that our proposed system with review similarity calculation using the frequency of the commonly used terms outperformed conventional CF system with 10% statistical significance level. Our study sheds a light on the application of users' review information for facilitating electronic commerce by recommending proper items to users.

Key Words : Recommender system, Collaborative filtering, Text mining, TF-IDF, App Store.

Received : May 20, 2015 Revised : June 16, 2015 Accepted : June 16, 2015
Type of Submission : Outstanding Conference Paper Corresponding Author : Hyunchul Ahn

저 자 소 개

전 병 국
원광대학교 정보전자상거래학부에서 학사학위를 취득하였으며, 현재 국민대학교 비즈니스IT전문대학원에서 비즈니스IT전공으로 석사과정에 재학 중이다. 주요 관심분야는 데이터마이닝, 텍스트마이닝, 빅데이터, 경영정보시스템이다.

안 현 철
현재 국민대학교 경영대학 경영정보학부 부교수로 재직 중이다. KAIST에서 산업경영학사를 취득하고, KAIST 테크노경영대학원에서 경영정보시스템을 전공하여 공학석사와 박사를 취득하였다. 주요 관심분야는 금융 및 고객관계관리 분야의 인공지능 응용, 정보시스템 수용과 관련한 행동 모형 등이다.