

사용자 간 신뢰관계 네트워크 분석을 활용한 협업 필터링 알고리즘의 예측 정확도 개선*

최슬비

국민대학교 비즈니스IT전문대학원
(seulbimon@kookmin.ac.kr)

곽기영

국민대학교 비즈니스IT전문대학원
(kykwahk@kookmin.ac.kr)

안현철

국민대학교 비즈니스IT전문대학원
(hcahn@kookmin.ac.kr)

협업 필터링(Collaborative Filtering)은 유용성과 정교성 면에서 가장 성공적인 추천 알고리즘으로 평가받으며 산업계나 학계에서 많이 활용 및 연구되고 있지만, 기본적으로 사용자들이 평가한 점수에만 기반하여 추천결과를 생성하는 한계점이 있다. 이에 본 연구는 사용자가 상품을 구매할 때 자신이 신뢰하는 타인의 추천을 더 적극적으로 수용할 것이라는 점에 착안하여, 사용자의 평점 외에 사용자 간 신뢰관계를 소셜네트워크분석으로 분석한 결과를 추가로 반영하는 추천 알고리즘들을 제안하였다. 구체적으로 본 연구에서는 소셜네트워크분석에서 네트워크 내의 중심적 위치를 나타내는 척도인 내향 및 외향 중심성을 활용하여 사용자 간 유사도를 산출하는 알고리즘들과 사용자 신뢰 네트워크를 탐색하여 추천 대상이 되는 사용자가 직접·간접적으로 신뢰하는 사용자의 평가점수를 보다 높게 반영하는 알고리즘을 제안한 뒤 그 성능을 비교해 보았다. 실제 데이터에 적용하여 분석한 결과, 사용자 신뢰 네트워크의 내향 중심성 지수를 조건 없이 적용한 경우에는 오히려 정확도의 감소만을 야기하는 것으로 나타났고, 일정 임계치 이상의 외향 중심성을 갖는 사용자에 한해 내향 중심성 지수를 고려한 추천 알고리즘은 전통적인 협업 필터링에 비해 약간의 정확도 개선이 이루어짐을 확인할 수 있었다. 아울러, 사용자 신뢰 네트워크를 기반으로 탐색하는 알고리즘이 가장 우수한 성능을 보이는 것을 알 수 있었으며, 전통적인 협업 필터링과 비교해서도 통계적으로 유의한 수준의 정확도의 개선이 이루어짐을 확인할 수 있었다.

주제어 : 추천시스템, 협업필터링, 소셜네트워크분석, 중심성, 신뢰관계 네트워크

논문접수일 : 2016년 8월 20일 논문수정일 : 2016년 9월 19일 게재확정일 : 2016년 9월 27일
원고유형 : 일반논문(급행) 교신저자 : 안현철

1. 서론

협업 필터링 추천 시스템은 상품에 대한 사용자들의 평가 데이터를 가지고 추천의 대상이 되는 사용자와 선호도가 유사한 즉, 평가 패턴이 비슷한 사용자들(이웃)을 식별한 후, 과거에 구매했거나 선호했던 상품들 중 추천 대상자에게

적합한 상품을 추천하는 시스템이다. 협업 필터링 기반의 추천 시스템을 구현하는 과정에서 가장 중요한 것은 추천 대상자와 유사한 선호도 패턴을 보이는 사용자가 누구인지를 정확하게 식별하는 것이다. 그러나 전통적인 협업 필터링에서 사용자 간 유사도는 구매여부, 클릭스트림(clickstream) 데이터 혹은 평점과 같은 정보들을

* 이 논문은 2014년 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임
(NRF-2014S1A5A2A03064791).

활용해 산출되어 왔다(Jeon and Ahn, 2015). 이와 같은 정량적(quantitative)인 정보는 연산 처리에 큰 어려움이 없다는 장점이 있지만, 사용자들의 선호도를 뚜렷이 대표하는 기준이라고 확실하게 담보하기 어려우며 이에 따라 의문이 제기되고 있다(Zhang et al., 2014).

인터넷의 사용이 활성화되고 보편화 되면서 전 연령대에서 상품을 구매하는 패턴이 온라인 쇼핑으로 변화하고 있다. 더불어 Web 2.0 시대의 도래와 함께, 인터넷 상에 자신의 의견을 상대방과 공유하는 참여중심의 인터넷 환경이 활성화되고 있다(Chen et al., 2007). 이러한 배경 가운데 기존 연구들에 따르면 사용자들은 일반적으로 본인과 친분 관계가 있는 사용자들의 추천을 더 신뢰하고 선호하는 것으로 알려져 있다(Kim and Ahn, 2010; Liu and Lee, 2010; Shinha and Swearingen, 2001). 이는 사용자 간의 친분 관계가 신뢰성을 제고하는 효과가 있기 때문인 것으로 판단된다(Kim and Kim, 2014).

본 연구에서 새롭게 제안하는 추천 알고리즘은 단순히 상품에 대한 사용자 평가점수의 유사성만을 고려한 것이 아니라, 사용자들 사이에서 신뢰하는 사용자를 나타내는 데이터를 활용하여, 사용자 간 신뢰관계까지 함께 고려할 수 있도록 설계되었다. 이 때, 신뢰관계 데이터를 분석하기 위해서 소셜네트워크분석 기술을 사용하였다.

본 논문의 구성은 다음과 같다. 우선 2장에서는 협업 필터링과 소셜네트워크분석의 기본 개념과 원리에 대해서 살펴본다. 아울러 소셜네트워크분석을 이용해 추천 시스템의 성능을 개선하고자 시도한 관련 기존 연구들에 대해서도 살펴본다. 이어 3장에서는 본 연구에서 제안하는 협업 필터링과 소셜네트워크분석을 이용한 새로

운 추천 알고리즘에 대해서 설명하고, 4장에서는 실증분석을 위한 데이터 소개, 실험 설계 및 실험 결과에 대하여 제시한다. 마지막 5장에서는 실험 결과에 따른 연구 성과 및 의의를 설명하고, 본 연구의 한계점 및 향후 연구 방향을 제시한다.

2. 이론적 배경

2.1 추천 시스템

추천 시스템은 사용자의 행동 살펴보기(Breese et al., 1998), Top-N 상품 추천 리스트 만들기 또는 사용자의 평가점수를 묻거나 예측하는 방법을 통해 그들이 흥미를 갖거나 구매하기 원하는 상품을 쉽게 찾도록 도와주는 데이터 분석 기술 기반의 정보 필터링(information filtering)시스템을 일컫는다(Sarwar et al., 2001). 추천 시스템의 주된 목적은 사용자와 상품에 대한 접근 가능한 다양한 정보를 분석하여 선호도가 높을 것으로 예측되는 상품을 추천하는 것이다(Lee and Park, 2007).

추천 시스템에 적용되는 알고리즘에는 크게 내용기반 필터링(Content-based filtering, CB)과 협업 필터링(Collaborative Filtering, CF)이 있다. 내용 기반 필터링은 상품 간의 유사성(item-to-item similarity)을 기반으로 하여 추천 결과를 생성하지만, 협업 필터링은 사용자 간 유사성(user-to-user similarity)에 따라 추천결과를 생성한다. 협업 필터링이 내용기반 필터링 보다 상대적으로 우수한 추천 정확도를 보이기 때문에 가장 대표적인 추천 기법으로 알려져 있다(Kim and Ahn, 2009; Kim and Ahn 2011).

2.2 협업 필터링

일반적으로 협업 필터링 알고리즘 기반 추천 프로세스는 다음과 같은 절차에 의해 사용자들 위한 추천 대상 상품을 결정한다(Herlocker et al., 1999).

단계 1. 사용자-상품 평가점수 행렬 구성

협업 필터링 추천 알고리즘의 첫 단계는 <Figure 1>과 같이 N 개의 상품에 대한 M 명 사용자의 평점을 $N \times M$ 의 사용자-상품 평가점수 행렬로 구성하는 것으로 시작된다.

	Item 1	Item 2	Item 3	...	Item N
User 1	5		4		2
User 2		1	4		3
User 3	2	2	3		2
...					
User $M-1$			2		5
User M	1	3	5		1

<Figure 1> Example of user-item rating matrix

단계 2. 사용자 간 유사도 계산

단계 2에서는 사용자 간의 유사도를 계산하여 선호도가 유사한 사용자를 탐색하는 작업이 수행된다. 유사도를 측정하는 방법으로는 주로 피어슨 상관계수(Pearson correlation coefficient, PCC) 혹은 코사인 유사도(Cosine similarity)가 사용된다. 이 중, PCC가 가장 많이 사용되며 식 (1)과 같이 계산된다.

$$S_{x,y} = \frac{\sum_i (R_{x,i} - \bar{R}_x) \cdot (R_{y,i} - \bar{R}_y)}{\sqrt{\sum_i (R_{x,i} - \bar{R}_x)^2} \cdot \sqrt{\sum_i (R_{y,i} - \bar{R}_y)^2}} \quad (1)$$

위의 식에서 $S_{x,y}$ 는 사용자 x 와 사용자 y 의

유사도이고, $R_{x,i}$ 는 사용자 x 의 상품 i 에 대한 평가점수이고, $R_{y,i}$ 는 사용자 y 의 상품 i 에 대한 평가점수이다. \bar{R}_x 은 사용자 x 의 평가점수 평균값이고, \bar{R}_y 은 사용자 y 의 평가점수 평균값이다.

단계 3. 이웃 선택

사용자 간의 유사도가 산출된 후, 단계 3에서는 이 유사도를 기반으로 추천 대상 사용자와 가장 유사한 N 명의 이웃을 선택하게 된다.

단계 4. 추천 상품 결정

마지막 4단계에서는 각 상품에 대한 고객의 선호도를 예측하여 최종 추천 상품을 결정하는 작업이 이루어진다. 특정 상품에 대한 추천 대상 사용자의 평가점수 예측은 다음의 식 (2)를 통해서 이루어진다. 이 때, $P_{x,i}$ 는 추천 대상자 x 의 상품 i 에 대한 평가점수를 나타낸다.

$$P_{x,i} = \bar{R}_x + \sum_{z \in N} (R_{z,i} - \bar{R}_z) \times \frac{S_{x,z}}{\sum_{z \in N} |S_{x,z}|} \quad (2)$$

위 식에서 \bar{R}_x 는 사용자 x 의 평가점수 평균값이고, $R_{z,i}$ 는 이웃 사용자 z 의 상품 i 에 대한 평가점수이다. $S_{x,z}$ 는 추천 대상자 x 와 이웃 사용자 z 사이의 유사도를 나타낸다. 그리고 N 은 유사한 사용자를 탐색하는 작업에서 선택된 가장 이웃하는 이웃 사용자들의 집합을 의미한다.

2.3 소셜네트워크분석과 연결정도 중심성

연결된 네트워크의 의미를 이해하는데 있어, 각 노드(node)의 개별적 특성보다는 전체적인 네트워크의 패턴을 분석하는데 초점을 맞추고 수

행되는 소셜네트워크분석은 오늘날 사회과학 뿐만 아니라 물리학, 의학, 마케팅 등 다양한 분야에서 많은 주목을 받고 있다(Park and Kwahk, 2013). 특히 컴퓨팅 기술이 발전하고 인터넷 사용이 보편화 되면서 대규모의 네트워크 데이터 분석이 용이해짐에 따라, 그 활용 범위는 더욱 광범위해져서 다양한 연구의 분석 기법으로 활용되고 있다(Kim and Im, 2014).

소셜네트워크분석에서 네트워크 구조를 파악하기 위한 척도로는 밀도, 호혜성, 이행성, 중심성, 구조적 공백 등이 사용된다. 이 중, 중심성은 어떤 개체가 전체 네트워크에서 중심에 위치하는 정도를 나타낸다. 구체적으로 중심성에는 연결정도 중심성(degree centrality), 근접 중심성(closeness centrality), 매개 중심성(betweenness centrality), 아이겐벡터 중심성(Eigenvector centrality) 등이 있다(Kwahk, 2014). 이 중에서 본 연구는 가장 대표적인 중심성 척도인 연결정도 중심성을 기반으로 연구를 진행하고자 한다.

연결정도 중심성은 액터가 네트워크 내에서 연결되어 있는 정도를 기반으로 측정된다. 가장 간단하게 중심성을 측정할 수 있는 방법으로, 한 액터와 직접 연결 관계를 맺고 있는 액터의 개수를 합하여 구할 수 있다(Jeong and Kim, 2013; Sohn, 2002). 연결정도 중심성 척도는 방향성을 고려해볼 수 있는데, 비방향 네트워크(nondirected network)는 내향·외향 연결정도 측정이 아닌, 단순히 연결 관계의 개수로 중심성을 평가한다. 하지만 방향 네트워크(directed network)에서는 내향·외향 연결정도 중심성을 구분하여 중심성을 평가할 수 있다. 일반적으로 외향 연결정도(outdegree)는 개방적 성향(expansiveness)을 나타내고 내향 연결정도(indegree)는 액터의 인기를 반영한다(Kwahk, 2014).

2.4 소셜네트워크분석과 추천 시스템

소셜네트워크 분석을 추천 시스템 연구에 반영하고자 하는 최초의 시도들은 단순히 소셜네트워크 정보를 추천 과정에 반영하는 것이 많았으며 본격적으로 사용하지는 않았다. Golbeck (2006)은 웹으로부터 수집한 소셜네트워크 정보를 기반으로 FilmTrust라는 새로운 영화 추천 시스템을 제안하였다. 이 연구에서는 평균과 비교하여 사용자들의 다양한 평가 데이터가 존재할 때 더 정확도가 높은 추천이 가능하다는 것을 보여주었다. Kim and Kim(2014)과 Liu and Lee (2010)는 친구관계 네트워크 정보를 이용하여 사용자와 친구관계인 사용자의 선호도 정보를 활용하거나 가중하는 새로운 추천 시스템을 제안하였다. 그들의 연구는 추천 대상자의 친구인 사용자의 선호도를 추천 과정에 더 많이 고려하였을 때 성능이 기존보다 더 개선되는 것을 확인할 수 있었다.

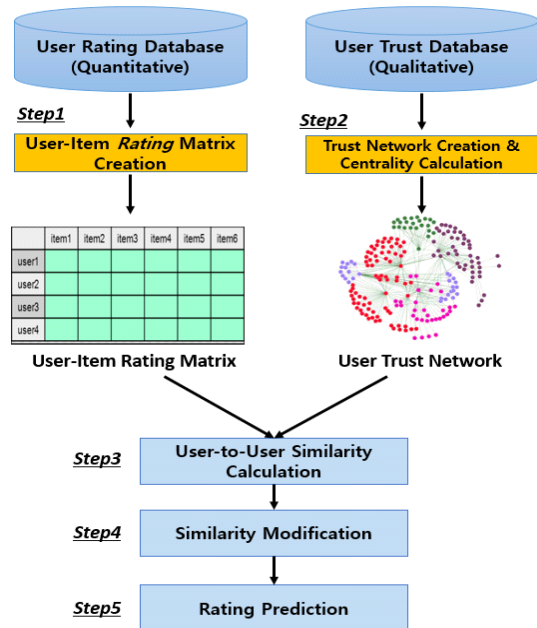
최근에는 소셜네트워크 정보를 탐색하는 데 그치는 것이 아니라, 소셜네트워크분석을 통한 결과물을 추천 과정에 이용하는 연구들이 진행되었다. Cho and Bang(2009)과 Park et al.(2009)은 모두 상품 추천 시스템에 소셜네트워크분석의 중심성 척도를 접목하였다. 다만 Cho and Bang(2009)은 동일한 사용자가 구매하는 상품들의 소셜네트워크를 활용한 반면, Park et al.(2009)은 동일한 상품을 구매하는 사용자들의 소셜네트워크를 활용했다는 점에서 서로 차이가 있다. 한편, Kim and Kim(2014)은 구조적 공백 분석을 실시하여 네트워크 내의 주요한 개체를 추출한 후 이들을 중심으로 한 군집을 형성하여 군집 색인 협업 필터링을 수행하였다. 실험 결과, 제안한 모형이 다른 비교 모형에 비해서 추천 성

과의 정확도가 가장 우수하였다.

앞서 언급된 연구들은 소셜네트워크분석과 협업 필터링을 통합한 추천 시스템을 제안하였다는 의의가 있다. 하지만, 외향 연결정도와 내향 연결정도의 의미를 갖는 방향성 네트워크를 분석하거나 사용자 간의 직접적인 신뢰관계를 나타내는 네트워크를 추천시스템에 활용한 사례는 기존 연구에서 찾아보기 어렵다.

3. 제안 알고리즘

본 연구에서는 소셜네트워크분석을 활용한 신뢰 네트워크 데이터를 추가로 반영하여, 전통적인 협업 필터링의 성능을 개선할 수 있는 새로운 추천 알고리즘을 제안한다. 본 연구의 제안 알고리즘은 <Figure 2>와 같이 총 5단계로 구성된다.



(Figure 2) Procedure of the proposed algorithm

단계 1. 사용자-상품 평가점수 행렬 도출

협업 필터링 기반 추천 시스템의 첫 번째 단계는 사용자들의 상품 평점 데이터들을 종합하여, 앞서 <Figure 1>에서 예시된 것과 같은 사용자-상품 평가점수 행렬을 도출하는 것이다.

단계 2. 신뢰관계 네트워크 구축 및 척도 계산

2단계에서는 사용자 간의 링크로 구성되어 신뢰관계를 표현하는 소셜네트워크를 만들고, 이로부터 중심성 척도를 계산한다. 본 연구에서는 신뢰관계 데이터가 방향·이진 네트워크로 나타나기 때문에 내향 연결정도와 외향 연결정도로 구분할 수 있다. 따라서 내향 연결정도를 기반으로 한 중심성과 외향 연결정도를 기반으로 한 중심성을 모두 계산한다.

단계 3. 사용자 간 유사도 산출

1단계에서 도출된 사용자-상품 평가점수 행렬을 참조하여 추천 대상자와 나머지 사용자들을 대상으로 평가점수 유사도를 산출한다. 본 연구에서는 평가점수 유사도 계산 방법으로 PCC를 사용한다. PCC 식은 앞서 2장에서 설명한 식 (1)에 의해 산출된다.

단계 4. 소셜네트워크분석을 통한 유사도 조정

이 단계에서는 2단계와 3단계에서 도출된 사용자 신뢰관계 네트워크 데이터와 사용자 간 유사도를 통합하여 전체적으로 사용자 간 유사도를 조정하게 되는데, 이를 반영한 사용자 간 유사도 $S^*_{x,y}$ 는 다음 식 (3)과 같이 산출된다.

$$S^*_{x,y} = amp_{x,y} \times S_{x,y} \quad (3)$$

위 식에서 $S_{x,y}^*$ 는 신뢰관계 네트워크 정보를 추가로 고려하여 사용자 x 와 사용자 y 의 조정된 유사도를 의미한다. $amp_{x,y}$ 는 사용자 x 와 사용자 y 사이의 유사도를 확대해주는 조정계수이다.

본 연구에서는 유사도 조정 작업 즉, $amp_{x,y}$ 를 도출하기 위해 크게 3가지 접근법을 제시한다. 첫 번째 접근법(Trust CF-All)은 사용자 y 의 신뢰관계 네트워크에 대한 내향 연결정도 중심성을 고려하여 유사도를 확대하는 것이다. 예를 들어, A 사용자의 내향 연결정도 중심성 값이 높다면, 이는 다른 사용자들로부터 높은 신뢰를 받고 있다는 의미이므로 다른 사용자들이 A의 추천을 적극적으로 수용할 가능성이 높다. 이러한 배경에서, 첫 번째 접근법에서는 $amp_{x,y}$ 를 다음의 식 (4)와 같이 적용한다.

$$amp_{x,y} = 1 + (IC_y)^\mu \quad (4)$$

상기 식에서 μ 는 승수이고, IC_y 는 사용자 y 의 내향 연결정도 중심성이다. 여기서 승수 μ 는 유사도에서 내향 연결정도 중심성을 비중 있게 반영하기 위해 결정해야하는 지표로써, 시행착오(trial-and-error)를 거쳐 최적의 값을 찾아야 한다.

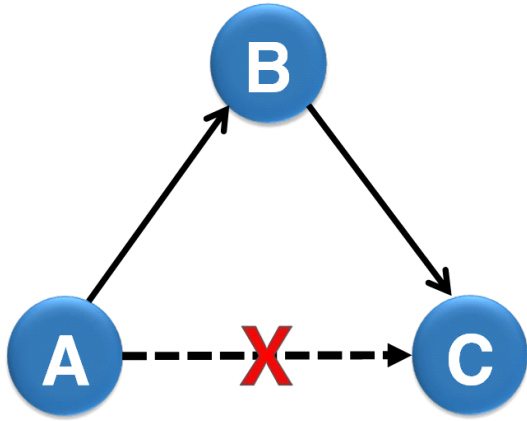
두 번째 접근법(Trust CF-Conditional)은 앞서 설명한 첫 번째 접근법과 유사하지만 이를 확장한 방법으로, 외향 연결정도 중심성이 특정 임계치(threshold) p 이상인 사용자 x 에 대해서만 다른 사용자 y 의 내향 연결정도 중심성을 고려하는 것이다. 예를 들어, B 사용자의 외향 연결정도 중심성의 값이 높다면, 이는 사용자 B가 다른 사람을 신뢰하는 성향이 강하다고 볼 수 있기 때문에 내향 연결정도 중심성의 값이 높은 사용자 A

의 추천을 긍정적으로 받아들일 수 있다. 반대로 B 사용자의 외향 중심성이 어떤 임계치 이하 수준으로 낮은 상황이라면, 이 사용자의 경우 타인에 대한 전반적인 신뢰가 낮은 상황이므로 타인의 신뢰 정도를 추가적으로 고려하지 않는 것이 더 나을 수 있다. 이러한 특징을 반영하여, 두 번째 접근법에서는 $amp_{x,y}$ 를 다음의 식 (5)와 같은 방식으로 계산하도록 하였다.

$$amp_{x,y} = \begin{cases} 1 & (OC_x < p) \\ 1 + (IC_y)^\mu & (OC_x \geq p) \end{cases} \quad (5)$$

위 식에서 p 는 모형 설계자가 결정할 특정 임계치이고, OC_x 는 사용자 x 의 외향 연결정도 중심성을 나타낸다.

마지막 세 번째 접근법(Trust CF-Search)은 2단계에서 구축된 신뢰 네트워크 데이터를 직접 탐색하여 유사도에 반영하는 방법이다. 이 접근법에서는 ‘직접적인 신뢰관계’와 ‘간접적인 신뢰관계’를 고려하는데, 예를 들어 “사용자 A는 사용자 B를 신뢰한다”고 응답했거나, “사용자 B가 사용자 C를 신뢰한다”고 응답한 경우, A와 B, B와 C는 직접적인 신뢰관계가 있는 것으로 정의한다. 하지만, <Figure 3>에 제시된 상황과 같이 만약 $A \rightarrow B$, $B \rightarrow C$ 간에는 직접 신뢰관계가 있지만, “사용자 A가 사용자 C를 신뢰한다”는 응답은 없는 상황이라면, 이 경우에도 비록 A는 C를 직접적으로 신뢰하지 않지만, A가 직접적으로 신뢰하는 사람인 B가 C를 직접적으로 신뢰하고 있기 때문에, 사용자 A는 중개자 B를 매개로 사용자 C와 간접적인 신뢰관계를 형성하고 있다고 볼 수 있다. 이러한 사용자 A와 C의 관계를 본 연구에서는 간접적인 신뢰관계로 정의한다.



〈Figure 3〉 Example of direct and indirect trust relationships

이러한 특징을 반영하여, 세 번째 접근법에서는 $amp_{x,y}$ 를 다음의 식 (6)와 같은 방식으로 계산한다.

$$amp_{x,y} = \begin{pmatrix} 1 & (not-trust) \\ \mu_1 & (indirect-trust) \\ \mu_2 & (direct-trust) \end{pmatrix} \quad (6)$$

위 식에서 μ_1 은 사용자 x 와 사용자 y 의 간접 신뢰관계 승수이고, μ_2 는 직접 신뢰관계 승수이다. 직접적인 신뢰관계의 경우, 간접적인 신뢰관계보다 더 높은 수준으로 가중치를 반영하는 것이 합리적이므로, $\mu_1 \leq \mu_2$ 가 항상 만족되어야 한다.

단계 5. 선호도 예측

마지막 5단계에서는 추천 대상자의 각 상품별 선호도(예상평점)를 예측하는 작업이 이루어진다. 각 상품별 선호도 예측은 앞서 2장에서 소개된 식 (2)를 변형한 식 (7)을 사용하여 계산한다.

$$p_{x,i} = \bar{r}_x + \frac{\sum_{y \in N} (r_{y,i} - \bar{r}_y) \times (S^*_{x,y})}{\sum_{y \in N} S^*_{x,y}} \quad (7)$$

이처럼 선호도 예측 작업이 끝나면, 예상 평가 점수가 높게 나온 상품들을 중심으로 추천 대상자가 아직 경험해보지 않은 상품들을 중심으로 추천 대상 상품을 결정한다(Jeon and Ahn, 2015).

4. 실증분석

4.1 실험 데이터 선정

본 연구에서 제안하는 추천 알고리즘들의 예측 정확도를 확인하기 위하여 LibRec으로부터 제공받은 데이터를 사용해 실험을 수행한다. 제공된 연구데이터는 사용자들의 영화 선호도 평가점수 자료와 사용자 간 신뢰관계 네트워크에 대한 자료이다(Golbeck, 2006). 영화 선호도 평가점수 자료는 사용자 ID, 영화 ID, 평점으로 구성되어 있으며, 총 35,497건의 평가 데이터이다. 사용자 간 신뢰관계 데이터는 신뢰하는 사용자 ID, 신뢰받는 사용자 ID로 구성되어 있으며, 방향/이진 네트워크 그래프로 표현이 가능하다.

본 연구에서 제안하는 추천 알고리즘에 대한 다양한 접근법들(Trust CF-All, Trust CF-Conditional, Trust CF-Search)의 유용성을 확인하기 위하여, 본 연구에서는 사용자 간 신뢰관계 네트워크 데이터에 UCINET 6를 이용하여 내향 및 외향 중심성 지수를 산출하고, Microsoft Excel VBA로 프로그램된 별도의 SW를 활용하여 사용자 간 직접 및 간접 신뢰관계를 나타내는 사용자-사용자 간 신뢰관계 행렬(User-to-User

Trust Matrix)을 산출하였다. 아울러, 신뢰관계를 가중하여 사용자 간 유사도를 산출하는 본 연구의 제안 알고리즘 역시 Microsoft Excel VBA로 실험용 소프트웨어를 구현하여 실험하였다.

제안 추천 알고리즘의 경우, (1) 조건없이 신뢰관계 네트워크의 내향 중심성을 고려하는 방식(Trust CF-All)과, (2) 외향 중심성이 특정 임계치 이상인 사용자에게 한하여 신뢰관계 네트워크의 내향 연결정도 중심성을 반영하는 방식(Trust CF-Conditional) 그리고 (3) 직접 신뢰 및 간접 신뢰관계를 적용하는 방식(Trust CF-Search)을 모두 적용해 본 후, 그 성능을 비교해 보고자 하였다. 그리고, 전통적인 CF 알고리즘을 적용하여 그 결과를 상기 세 접근법의 벤치마크 대상으로 사용하였다.

4.2 실험 결과

본 연구에서는 사용자의 평가점수가 입력된 영화에 대해서 제안한 추천 알고리즘으로 예상 평가점수를 도출한 다음, 실제 평가점수와 비교했을 때 평균 오차가 가장 적은 추천 알고리즘이 무엇인지를 확인해보는 방식으로 검증을 진행하였다(Kim and Kim 2014). 이 때, 점수 간의 오차는 추천 시스템 관련 연구에서 가장 많이 사용되는 척도인 평균 MAE(Mean Absolute Error)를 활용하였다(Breese et al., 1998; Sarwar et al., 2001).

우선 첫 번째로 제시한 조건없이 신뢰관계 네트워크의 내향 중심성을 고려하는 접근법(Trust CF-All)의 결과는 다음의 <Table 1>과 같다. 이 표에서 볼 수 있듯이, Trust CF-All은 유감스럽게도 전통적인 협업 필터링 보다 성능의 개선을 가져오지 못함을 확인하였다. 특히 승수의 값이 커질수록 성능이 더 나빠지는 것으로 나타나, 신뢰

관계 네트워크의 내향 중심성 지수를 더 강하게 고려할수록 오히려 추천 알고리즘의 성능에 악영향을 끼침을 알 수 있었다. 즉, 상대적으로 사용자들 사이에서 신뢰도가 높은 사용자의 의견을 추천 과정에 반영한다는 점에서 성능의 개선을 기대했지만, 결과적으로 성능 개선 효과는 없는 것으로 분석되었다.

<Table 1> Predictive accuracy of Trust CF-All

Approach	Multiplying Coefficient(μ)	Average MAE
Conventional CF (Comparison)	0	0.564966
Trust CF-All	1	0.565134
	2	0.565212
	3	0.565792
	4	0.568665
	5	0.580273

한편 두 번째로 제시한 외향 중심성이 특정 임계치 이상인 사용자에게 한하여 신뢰관계 네트워크의 내향 연결정도 중심성을 반영하는 접근법(Trust CF-Conditional)의 실험결과가 다음의 <Table 2>에 제시되어 있다.

이 표에서 볼 수 있듯이, 외향 중심성을 기준으로 적절하게 임계치를 설정하고, 해당 임계치 이상의 중심성을 보인 사용자에게 한하여 내향 중심성을 가중 반영할 경우 예측정확도가 소폭 상승함을 알 수 있다. 본 실험에서 적용한 여러 설정값들 중에서는 p 가 0.1이고, μ 가 2일 때, 평균 MAE가 0.564909가 되어 가장 높은 예측 정확도가 산출되었다. 이는 전통적인 CF의 평균 MAE에 비해 5.7×10^{-5} 감소한 수치이다. 이러한 실

<Table 2> Predictive accuracy of Trust CF-Conditional

Threshold for Outdegree Centrality(p)	Multiplying Coefficient for Indegree Centrality(μ)	Average MAE
0.5	1	0.564966
	2	0.564968
	3	0.564987
0.45	1	0.564961
	2	0.564958
	3	0.564967
0.4	1	0.564961
	2	0.56496
	3	0.564974
0.35	1	0.564956
	2	0.564952
	3	0.564990
0.3	1	0.564948
	2	0.564943
	3	0.564981
0.25	1	0.564948
	2	0.564943
	3	0.564984
0.2	1	0.564947
	2	0.564946
	3	0.565005
0.15	1	0.564946
	2	0.564927
	3	0.564996
0.1	1	0.564959
	2	0.564909
	3	0.564927
0.05	1	0.564972
	2	0.564962
	3	0.565114

험의 결과는 사용자 신뢰관계의 개방적 성향(일종의 사용자가 갖는 신뢰에 대한 본원적 성향)이 높아야만 다른 사용자의 신뢰관계 네트워크에

대한 내향 중심성을 추천 과정에서 고려하는 것이 의미가 있다는 점을 시사한다.

<Table 3>에는 직접 신뢰 및 간접 신뢰관계를 적용하는 방식(Trust CF-Search)의 실험결과가 제시되어 있다. 이 표의 결과를 보면, <Table 1>이나 <Table 2>에 제시된 다른 접근법들과 비교해 Trust CF-Search 방식이 월등히 우수한 예측 정확도를 산출함을 알 수 있다. 특히 승수값이 커질수록 예측정확도가 상승하는 패턴을 보여, $\mu^1 = 5$, $\mu^2 = 7$ 일 때 가장 낮은 평균 MAE (0.564846)를 나타내고 있음을 알 수 있는데, 이는 직·간접적인 신뢰 관계를 추천과정에서 고려하는 것이 중심성을 고려하는 다른 접근법에 비해 훨씬 효과적이라는 사실을 우리에게 보여주고 있다.

<Table 3> Predictive accuracy of Trust CF-Search

Approach	Multiplying Coefficient(μ)	Average MAE
Trust CF-Search	$\mu^1 = 2$ $\mu^2 = 3$	0.564912
	$\mu^1 = 3$ $\mu^2 = 5$	0.564871
	$\mu^1 = 5$ $\mu^2 = 7$	0.564846

이상 소개한 세 가지 접근법의 예측 정확도와 비교 기법인 전통적인 CF의 예측 정확도를 종합적으로 정리한 결과가 다음의 <Table 4>에 제시되어 있다. 이 표를 통해서 알 수 있듯이, Trust CF-All < 전통적인 CF < Trust CF-Conditional < Trust CF-Search 순으로 예측 정확도가 산출됨을 알 수 있다.

(Table 4) Summary of all the experimental results

Recommendation Algorithms		Average MAE	Optimal Setting
Conventional CF		0.564966	-
Proposed Algorithm	Trust CF-All	0.565134	$\mu = 1$
	Trust CF-Conditional	0.564909	$\mu = 2$ $OC_x \geq 0.1$
	Trust CF-Search	0.564846	$\mu^1 = 5$ $\mu^2 = 7$

끝으로 이러한 성과의 차이가 과연 통계적으로 유의한지를 검증하기 위해, 귀무가설 H_0 는 $\mu_A = \mu_B$, 대립가설 H_a 는 $\mu_A > \mu_B$ (μ_X : 기법 X의 평균 MAE)로 하는 대응표본 t-검정을 수행하였다. 대응표본 t-검정에서 <Table 5>는 대응표본 t-검정의 결과를 나타내고 있다. 이 표에서 볼 수 있듯이, 제안 알고리즘 중 가장 우수한 성능을 보인 Trust CF-Search는 전통적인 CF와 90% 신뢰수준, Trust CF-All과 99% 신뢰수준 하에서 통계적으로 유의한 성과 차이를 보이는 것으로 나타났다. 하지만, Trust CF-Conditional과는 그 차이가 통계적으로 유의하지 못하였다. Trust CF-Conditional은 전통적인 CF와도 통계적으로 유의한 수준의 성능 차이를 나타내지 못하고 있는 것으로 나타났다. 결과적으로, 제안된 3가지

(Table 5) t-values of the paired samples t-test

	Trust CF-All	Trust CF-Conditional	Trust CF-Search
Conventional CF	-4.270**	1.026	1.495*
Trust CF-All		4.562**	3.652**
Trust CF-Conditional			0.809

* statistically significant at 10%,

** statistically significant at 1%

알고리즘 중에서는 오로지 Trust CF-Search만 전통적인 CF와 비교해 다소 유의한 성과 개선이 이루어짐을 확인할 수 있었다.

5. 결론

본 연구에서는 사용자들의 영화 평점만을 활용하여 추천 결과를 생성하는 기존의 협업 필터링을 개선하기 위해, 신뢰 기반 사용자 네트워크를 추가로 고려하여 협업 필터링의 성능 높일 수 있는 새로운 추천 알고리즘을 제안하였다. 구체적으로, 사용자의 신뢰관계 네트워크에서 내향 연결정도 중심성을 고려하여 유사도를 확대하거나 혹은 외향 연결정도 중심성이 특정 임계치 이상인 사용자에 한하여 이웃 사용자의 내향 연결정도 중심성을 활용하는 방법, 마지막으로 네트워크를 직접적으로 탐색하여 직접적 및 간접적 신뢰관계를 반영하는 3가지 접근법을 제시하였다. 아울러, 이 3가지 접근법 중 어느 것이 더 우수한 예측 정확도를 보이는지 실증분석을 통해 확인하였다. 그 결과, 두 번째와 세 번째 접근법이 전통적인 CF에 비해 더 우수한 예측 정확도를 보인다는 사실과 그 중에서도 세 번째 접근법이 가장 우수한 추천 성능을 보인다는 사실을 확인할 수 있었다. 그리고 이를 통해 소셜네트워크 데이터를 추천알고리즘에 반영할 경우 사용자 간의 신뢰관계를 직접적으로 탐색한 결과를 고려하는 것이 예측의 효과성을 제고하는데 있어 가장 좋은 대안임을 제시하였다.

대다수의 추천 시스템 연구에서 사용되어 온 평점 데이터에만 의존하지 않고, 오늘날 중요성이 크게 대두되고 있는 사용자 간 신뢰관계 네트워크 데이터를 활용하여 예측 성능을 제고하는

새로운 접근법을 제시하였다는 점은 기존 연구와 본 연구가 크게 차별화되는 부분이라고 할 수 있다. 특히 기존 연구에서 시도되지 않았던 내·외향 중심성을 차별적으로 적용하는 방법론을 새롭게 제안하여 그 가능성을 확인한 점과 직·간접 신뢰관계를 반영하여 유의미하게 기존 CF의 예측 정확도를 개선할 수 있는 방안을 제시한 점은 본 연구가 갖는 중요한 학술적 의의라 생각된다.

하지만, 본 연구에서 예측 정확도를 유의미하게 개선하는 것으로 나타난 ‘Trust CF-Search’의 경우, 사용자간 유사도를 계산할 때 마다 매번 사용자 신뢰관계 네트워크를 탐색해야 하므로 다른 접근법에 비해 훨씬 많은 컴퓨팅 자원을 요구하게 된다. 따라서 Trust CF-Search의 효율성을 개선하기 위한 후속 연구가 추후 이루어져야 할 것으로 보인다. 또한, Trust CF-Search에서는 현재 중개인이 1명인 경우에 한해서만 간접 신뢰관계를 정의하여 사용하고 있는데, 중개인이 2명 이상인 간접 신뢰관계까지 고려하면 더 높은 예측 정확도를 얻을 가능성이 있다. 때문에 이러한 부분에 대한 확장 연구 역시 추후 필요할 것으로 판단된다. 끝으로 현재 제안된 연구 모형에는 임계치(p), 승수(μ) 등 실험자가 임의로 설정해야 할 변수들이 포함되어 있다. 이러한 변수들의 값을 최적화하기 위한 방법 역시 의미 있는 후속 연구의 주제가 될 것으로 예상된다.

참고문헌(References)

- Breese, J. S., D. Heckerman and C. Kadie, “Empirical analysis of predictive algorithms for collaborative filtering,” Proceedings of the 14th 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.
- Cho, Y. and J. Bang, “Social network analysis for new product recommendation,” *Journal of Intelligence and Information Systems*, Vol.15, No.4(2009), 183~200.
- Golbeck, J., “Generating predictive movie recommendations from trust in social networks,” Proceedings of the 4th International Conference on Trust Management, Lecture Notes in Computer Science, Vol. 3986, 2006, 93~104.
- Herlocker, J., J. Konstan, A. Borchers and J. Riedl, “An Algorithm Framework for Performing Collaborative Filtering,” Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in information Retrieval, 1999, 230~237.
- Jeon, B. K. and H. Ahn, “A Collaborative Filtering System Combined with Users’ Review Mining: Application to the Recommendation of Smartphone Apps,” *Journal of Intelligence and Information Systems*, Vol.21, No.2(2015), 1~18.
- Jeong, J. -H. and J. -W. Kim, “Collaborative Filtering Techniques Using Social Network Analysis for UCC Recommendation,” *Journal of Intelligence and Information Systems*, Vol.11, No.1(2013), 185~195.
- Kim, K. -J. and H. Ahn, “Hybrid recommender systems using cluster-indexing collaborative filtering and social network analysis,”

- Proceedings of the 2010 Fall Conference of the Korean Society of Management Information Systems, 2010, 604-609.
- Kim, K. -J. and H. Ahn, "User-Item Matrix Reduction Technique for Personalized Recommender Systems," *Journal of Information Technology Application & 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. G. 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.
- Kim, M. S. and I. Im, "Resolving the 'Gray sheep' Problem Using Social Network Analysis (SNA) in Collaborative Filtering (CF) Recommender Systems," *Journal of Intelligence and Information System*, Vol.20, No.2(2014), 137~148.
- Kwahk, K. Y., *Social Network Analysis*, Cheongram, 2014.
- Lee, J. S. and S. D. Park, "Performance Improvement of a Movie Recommendation System using Genre-wise Collaborative Filtering," *Journal of Intelligence and Information Systems*, Vol.13, No.4(2007), 65~78.
- Liu, F. and H. J. Lee, "Use of social network information to enhance collaborative filtering performance," *Expert Systems with Applications*, Vol.37, No.7(2010), 4772~4778.
- Park, J. H., Y. H. Cho and J. K. Kim, "Social network: a novel approach to new customer recommendations," *Journal of Intelligence and Information Systems*, Vol.15, No.1(2009), 123~140.
- Park, J. H. and K. -Y. Kwahk, "The Effect of Patent Citation Relationship on Business Performance: A Social Network Analysis Perspective," *Journal of Intelligence and Information Systems*, Vol.19, No.3(2013), 127~139.
- Sarwar, B., G. Karypis, J. Konstan and J. Riedl, "Item-based collaborative filtering recommendation algorithms," *Proceeding of the 10th International Conference on World Wide Web*, 2001, 285~295.
- Shinha, R. and K. Swearingen, "Comparing recommendations made by online systems and friends," *Proceedings of the DELOS-NSF Workshop on Personalization and Recommender Systems in Digital Libraries*, 2001.
- Sohn, D. W., *Social Network Analysis*, Kyungmoon Publishing, 2002.
- Zhang, Z., D. Zhang and J. Lai, "urCF: User Review Enhanced Collaborative Filtering," *Proceedings of the 20th Americas Conference on Information Systems*, 2014.

Abstract

Enhancing Predictive Accuracy of Collaborative Filtering Algorithms using the Network Analysis of Trust Relationship among Users*

Seulbi Choi**·Kee-Young Kwahk***·Hyunchul Ahn****

Among the techniques for recommendation, collaborative filtering (CF) is commonly recognized to be the most effective for implementing recommender systems. Until now, CF has been popularly studied and adopted in both academic and real-world applications. The basic idea of CF is to create recommendation results by finding correlations between users of a recommendation system. CF system compares users based on how similar they are, and recommend products to users by using other like-minded people's results of evaluation for each product. Thus, it is very important to compute evaluation similarities among users in CF because the recommendation quality depends on it. Typical CF uses user's explicit numeric ratings of items (i.e. quantitative information) when computing the similarities among users in CF. In other words, user's numeric ratings have been a sole source of user preference information in traditional CF. However, user ratings are unable to fully reflect user's actual preferences from time to time. According to several studies, users may more actively accommodate recommendation of reliable others when purchasing goods. Thus, trust relationship can be regarded as the informative source for identifying user's preference with accuracy.

Under this background, we propose a new hybrid recommender system that fuses CF and social network analysis (SNA). The proposed system adopts the recommendation algorithm that additionally reflect the result analyzed by SNA. In detail, our proposed system is based on conventional memory-based CF, but it is designed to use both user's numeric ratings and trust relationship information between users when calculating user similarities. For this, our system creates and uses not only user-item rating matrix, but also

* This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government(NRF-2014S1A5A2A03064791).

** Master's Candidate, Graduate School of Business IT, Kookmin University

*** Professor, Graduate School of Business IT, Kookmin University

**** Corresponding Author: Hyunchul Ahn

Graduate School of Business IT, Kookmin University

77 Jeongneung-ro, Seongbuk-gu, Seoul 02707, Korea

Tel: +82-2-910-4577, Fax: +82-2-910-4017, E-mail: hcahn@kookmin.ac.kr

user-to-user trust network. As the methods for calculating user similarity between users, we proposed two alternatives – one is algorithm calculating the degree of similarity between users by utilizing in-degree and out-degree centrality, which are the indices representing the central location in the social network. We named these approaches as ‘Trust CF – All’ and ‘Trust CF – Conditional’. The other alternative is the algorithm reflecting a neighbor’s score higher when a target user trusts the neighbor directly or indirectly. The direct or indirect trust relationship can be identified by searching trust network of users. In this study, we call this approach ‘Trust CF – Search’.

To validate the applicability of the proposed system, we used experimental data provided by LibRec that crawled from the entire FilmTrust website. It consists of ratings of movies and trust relationship network indicating who to trust between users. The experimental system was implemented using Microsoft Visual Basic for Applications (VBA) and UCINET 6. To examine the effectiveness of the proposed system, we compared the performance of our proposed method with one of conventional CF system. The performances of recommender system were evaluated by using average MAE (mean absolute error). The analysis results confirmed that in case of applying without conditions the in-degree centrality index of trusted network of users(i.e. Trust CF - All), the accuracy (MAE = 0.565134) was lower than conventional CF (MAE = 0.564966). And, in case of applying the in-degree centrality index only to the users with the out-degree centrality above a certain threshold value(i.e. Trust CF - Conditional), the proposed system improved the accuracy a little (MAE = 0.564909) compared to traditional CF. However, the algorithm searching based on the trusted network of users (i.e. Trust CF – Search) was found to show the best performance (MAE = 0.564846). And the result from paired samples t-test presented that Trust CF - Search outperformed conventional CF with 10% statistical significance level. Our study sheds a light on the application of user’s trust relationship network information for facilitating electronic commerce by recommending proper items to users.

Key Words : Recommender system, Collaborative filtering, Social network analysis, Centrality, Trust relationship network

Received : August 20, 2016 Revised : September 19, 2016 Accepted : September 27, 2016
Publication Type : Regular Paper(Fast-track) Corresponding Author : Hyunchul Ahn

저 자 소개



최슬비

현재 국민대학교 비즈니스IT전문대학원에서 석사과정에 재학 중이다. 국민대학교 경영정보학 학사 학위를 취득하였으며, 주요 관심분야는 추천 시스템, CRM이다.



박기영

현재 국민대학교 경영대학 및 비즈니스IT전문대학원 교수로 재직 중이다. 서울대학교 경영대학을 졸업하고 KAIST 경영과학과 및 테크노경영대학원에서 석사 및 박사학위를 취득하였다. 주요 연구관심 분야는 Social network analysis and its application, Data analytics, Knowledge discovery and management, IT-enabled organizational agility 등이다.



안현철

현재 국민대학교 경영대학 및 비즈니스IT전문대학원 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.