



The Study of Recommendation Algorithm's Predictive Accuracy Improvement Using Structural Holes on Trust-based Social Networks

Boo-Sik Kang^{*}

Division of Service Management, MokWon University

ABSTRACT

Improving predictive accuracy of recommendation algorithms is a major work in the area of recommender systems. Collaborative filtering is the most popular method for product recommender systems. User-based collaborative filtering recommends products using the information about product preference of Neighbors. Recently, some studies enhancing predictive accuracy of recommender systems using information about the relationship of friend or trust between users has been published. This study proposed a method constructing trust-based social networks and applying structural holes for improving predictive accuracy of collaborative filtering. It constructs a social network using dataset represented trust relationship and finds user impact using structural holes that is one of methods for social network analysis. Neighbor's similarities of a target user for recommendation are adjusted by neighbor's impact which was found in before procedure. This study experimented two techniques for adjusting similarities. LinearImpCF adjusts neighbor's similarities multiplying by α and neighbor's impact. ExpImpCE adjusts neighbor's similarities multiplying neighbor's impact to the power α . To validate, the proposed algorithms were applied to filmtrust dataset. The results of 10-fold cross validation showed that mean MAEs of LinearImpCF and ExpImpCF were lower than mean MAE of conventional CF. We knew that the proposed method improved the predictive accuracy slightly. To test statistical significance, we experimented 10-fold cross validation repeatedly three times. We confirmed the statistical significance by paired t-test using experiment results. In conclusion, we knew that the proposed recommendation algorithm combined collaborative filtering and user's impact by structural holes on trust-base social networks between users improved the predictive accuracy.

© 2017 KKITS All rights reserved

KEYWORDS: Predictive accuracy, Collaborative filtering, Structural holes, Social network analysis, Trust-base social networks, Recommender systems, Recommendation algorithm

ARTICLE INFO: Received 25 January 2017, Revised 7 February 2017, Accepted 10 February 2017.

^{*}Corresponding author is with the Division of Service Management, MokWon University, 88 Doanbuk-ro Seo-gu

Daejeon, 35349, KOREA.

E-mail address: bookang@mokwon.ac.kr

1. 서론

협업필터링은 상품추천시스템에서 가장 많이 활용되고 있는 추천방식이다[1,2,3]. 협업필터링은 크게 상품기반 협업필터링과 사용자기반 협업필터링으로 구분할 수 있다. 상품기반 추천방식은 사용자가 구매했거나 좋아하는 상품과 유사한 상품을 추천하는 방식이다[1]. 사용자기반 추천방식은 사용자간의 유사성을 측정하여 유사한 특성을 갖는 이웃 사용자들의 구매 및 선호정보를 이용하여 추천하는 방식이다[2]. 협업필터링 추천시스템의 추천 대상 상품의 예측 정확도를 높이기 위한 주요 과정 중의 하나는 좀 더 정확하게 유사성을 찾아 반영하는 것이다. 지금까지는 주로 사용자의 상품 구매 여부, 구매 상품에 대한 사용자의 평점 등 정량화된 정보를 이용하여 유사성을 측정하였다.

최근 들어서는 웹에서 수집한 사회연결망 정보를 기반으로 추천 성능을 높이기 위한 연구[4], 친구관계 선호도 정보를 추천에 활용하여 추천 성능을 높이기 위한 연구[5] 등이 이루어지고 있다. 이들 연구들은 유사성을 측정하는 과정에서 사회연결망 정보, 친구 선호도 정보를 활용하는 것이 추천 정확도를 높이는 데 도움이 됨을 보여주고 있다. 또한, 사용자들 사이에서 신뢰관계가 형성된 경우 이를 활용할 경우 추천시스템의 추천 대상 상품의 예측 정확도를 높일 수 있음을 보이고 있다[6].

일상생활에서의 인터넷 활용이 활발해짐에 따라 많은 사용자 데이터가 축적되어 가고 있다. 사용자간에는 인터넷 상에서 자신의 의견을 상대방과 공유하는 참여중심의 사용자 환경도 증가하고 있다[7]. 사용자간의 사회적 연결정보를 사회연결망으로 구성하여 분석하면 이들 간의 관계를 파악할 수 있고[8], 사용자간의 신뢰관계를 활용하면 상품 추천 알고리즘을 개선시킬 수 있다[4,6].

선행연구에서 추천시스템의 예측 성능을 높이기 위해서 사용하는 사회연결망 분석 기법은 중심성[6]과 구조적 공백[5]이 사용되며, 중심성 활용연구는 많은 반면 구조적 공백에 대한 연구는 희박한 편이다. 또한 사회적 관계 자료의 수집이 현실적으로 어려움이 많아 실험실 환경의 자료수집으로 연구가 이루어지고 있다[5].

본 연구에서는 사용자간의 신뢰관계를 활용하여, 사용자기반 협업필터링의 추천 대상 상품의 예측 정확도를 높이는 방안을 제시한다. 이 방안에서는 사용자간의 신뢰관계 정보를 활용하여 이를 사회연결망으로 표현하고, 사회연결망의 구조적 공백 기법을 적용하여 신뢰 네트워크에 기반한 사용자 영향도를 추출하고, 사용자 영향도를 사용자간 유사성을 계산하는 과정에 반영함으로써 추천 대상 상품의 예측 정확도를 높이고자 한다.

본 논문의 구성은 다음과 같다. 제 2장에서는 사용자기반 협업필터링 방식과 사회연결망의 구조적 공백기법에 대해서 살펴본다. 제 3장에서는 본 연구에서 제안한 추천 알고리즘에 대해 설명한다. 제 4장에서는 제안한 기법에 filmtrust 데이터[9]를 적용한 실험을 하고 분석한다. 제 5장에서는 결론을 기술한다.

2. 협업필터링과 사회연결망 분석

2.1 사용자기반 협업필터링

사용자기반 협업필터링 방식은 추천 대상자와 성향이 유사한 이웃사용자들을 찾아내고, 이웃사용자들의 구매정보를 활용하여 추천 대상자가 구매하지 않은 상품들 중에서 가장 구매가능성이 높게 예측되는 상품들을 추천한다. 일반적으로 사용자기반 협업필터링 추천 프로세스는 사용자 유사도 계산, 이웃사용자 선정, 상품별 선호도 평가점수 예

측 및 상품추천으로 구성된다[10].

1단계에서는 사용자간 유사도를 계산한다. 유사도를 측정하는 방법으로는 피어슨 상관관계수, 코사인 유사도가 사용되는데, 일반적으로 피어슨 상관관계수를 가장 많이 사용되며 식(1)과 같이 계산된다.

$$s(x, y) = \frac{\sum_{j=1}^n (v_{xj} - \bar{v}_x)(v_{yj} - \bar{v}_y)}{\sqrt{\sum_{j=1}^n (v_{xj} - \bar{v}_x)^2} \sqrt{\sum_{j=1}^n (v_{yj} - \bar{v}_y)^2}} \quad (1)$$

식(1)에서 $s(x, y)$ 는 사용자 x 와 y 의 유사도를 나타내고, v_{xj} 와 v_{yj} 는 사용자 x 와 y 가 상품 j 에 부여한 선호도이며, \bar{v}_x 와 \bar{v}_y 는 사용자 x 와 y 의 평균선호도를 의미한다.

2단계에서는 추천 대상자와 유사도가 높은 N 명의 이웃사용자를 선정한다.

3단계에서는 추천 대상자를 위한 상품별 선호도 평가점수를 식(2)와 같이 계산한다.

$$p(x, i) = \bar{v}_x + \frac{\sum_{k \in N} s(x, k) \times (v_{ki} - \bar{v}_k)}{\sum_{k \in N} |s(x, k)|} \quad (2)$$

식(2)에서 $p(x, i)$ 는 추천 대상자 x 를 위한 추천 대상 상품 i 의 예측 평가점수이다. N 은 2단계에서 선정한 이웃사용자 집합을 나타낸다.

4단계에서는 추천 대상자 a 에게 선호도 예측 평가점수가 높은 상품순으로 추천 한다.

2.2 사회연결망 및 구조적 공백

사회연결망(Social Network) 분석을 위해서는 행위자들의 관계에 대한 데이터가 필요하며, 관계를

이용하여 네트워크를 구성한다[11].

사회연결망분석에서 관계 및 구조적 특성을 분석하기 위해 널리 사용되고 있는 기법으로 중심성(centrality), 밀도(density), 구조적 공백(Structural hole), 중심화(centralization) 등이 있다[8,11]. 밀도와 중심화는 주로 연결망 전체를 대상으로 하는 척도이며, 연결망내 각 노드에 대한 분석은 주로 중심성과 구조적 공백을 활용하여 분석하고 있다 [12,13,14,15].

중심성은 사회연결망 내에서 한 주체가 중심에 위치하는 정도를 나타내는 척도로 연결정도 중심성(degree centrality), 근접 중심성(closeness centrality), 사이 중심성(betweenness centrality), 위세 중심성(prestige centrality)이 있다[8,11]. 연결정도 중심성은 한 노드에 직접적으로 연결되어 있는 노드들의 관계 값의 합으로 측정할 수 있는데, 이는 연결망에서의 각 노드의 중심성을 나타낸다[8]. 중심성을 이용하여 협업필터링 방식의 적용이 어려운 신규고객, 신상품을 추천에 활용할 수 있다[12,13].

중심성이 연결망 내에서 한 고객이 얼마나 많은 고객과 관계를 맺고 있는가를 측정하는 척도라면 구조적 공백(structural holes)은 한 고객이 얼마나 다양한 고객들과 비중복적인 관계를 맺을 수 있는 위치에 있는가와 관련된 개념이다[8,11].

구조적 공백 개념은 Burt[16]가 제시한 것으로, Burt는 연결망에서의 유리한 위치를 비중복성과 낮은 제약성이라는 개념으로 측정하려고 했다[8]. 연결망내의 개체 i 와 j 의 제약성 c_{ij} 는 식(3)과 같이 계산한다[16].

$$c_{ij} = \left(p_{ij} + \sum_{k, k \neq i, k \neq j} p_{ik} p_{kj} \right)^2 \quad (3)$$

$$p_{ij} = \frac{a_{ij} + a_{ji}}{\sum_k (a_{ik} + a_{ki})} \quad (4)$$

p_{ij} 는 개체 i 가 연결망의 모든 개체와 갖는 관계의 총합에 대해 개체 i 가 개체 j 와 갖는 관계가 차지하는 비율을 나타낸다. a_{ij} 는 개체 i 로부터 개체 j 로 향하는 링크의 값을 나타낸다. 방향성이 없는 링크의 경우에는 $a_{ij}=a_{ji}$ 이다. 식(3)에서 제약성이 커지는 이유는 크게 2가지가 있다[8]. 개체 i 가 개체 j 에 투자하는 비율의 정도와 개체 i 가 제 3자 개체인 k 에게 투자를 많이 하는데, 3자 개체인 k 가 개체 j 에게 투자를 많이 할수록 제약성이 커지게 된다. 한 연결망에서 개체 i 의 제약성의 총합은 식(5)과 같이 구할 수 있다[16]. 만약 i 가 고립개체인 경우에는 $C_i = 1$ 이 된다.

$$C_i = \sum_j c_{ij} \tag{5}$$

개체 i 가 얼마나 구조적 공백의 위치를 잘 활용하고 있는 지를 나타내면 식(6)과 같이 정의할 수 있다[14,15].

$$imp(i) = 1/C_i \tag{6}$$

최근에는 협업필터링의 추천 정확도를 높이는 데에도 사회연결망이 활용되고 있다. 신뢰관계가 있는 사용자의 연결정도 중심성을 활용하여 사용자 기반 협업필터링의 추천 성능이 향상될 수 있음을 보이고 있다[6].

본 연구에서는 협업필터링의 예측 정확도를 높이는 데 사용자간 신뢰관계 기반 사회연결망의 구조적 공백 기법이 활용될 수 있음을 보이고자 한다.

3. 제안 추천 알고리즘

본 연구에서는 사용자 기반 협업필터링의 예측

정확도를 높이기 위해서 사용자간 신뢰관계를 활용하는 방안을 제시한다. 신뢰관계를 사회연결망으로 표현하고 구조적 공백 기법을 사용하여 각 사용자의 사회연결망에서의 영향도(식(6)참조)를 찾아내고 이를 협업필터링에서 활용함으로써 추천 대상 상품의 예측 정확도를 높이고자 하며, 이를 위한 절차는 다음 <그림 1>과 같다.

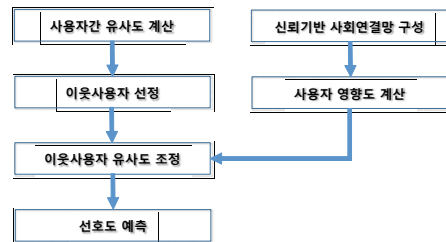


그림 1. 추천 알고리즘 절차
Figure 1. Procedure of the recommendation algorithm

사용자간 유사도 계산 단계에서는 사용자들의 상품 선호 데이터를 가지고 식(1)을 이용하여 사용자 x 와 y 간의 유사도 $s(x,y)$ 를 계산한다.

이웃사용자 선정단계에서는 추천 대상자와 선호도 유사도가 높은 순서로 N 명을 선정한다.

신뢰기반 사회연결망 구성 단계에서는 사용자간의 신뢰관계를 바탕으로 사회연결망을 구성한다. 신뢰관계 데이터에는 사용자 x 가 사용자 y 에 대한 신뢰여부가 표시되어 있다.

사용자 영향도 계산단계에서는 앞 단계에서 구성된 신뢰기반 사회연결망에서 구조적 공백기법을 사용하여 각 노드(사용자)의 제약값을 구하고(식(5)참조), 식(6)을 이용하여 영향도를 계산한다. 사회연결망에서 신뢰관계가 전혀 없는 경우 노드(사용자)의 영향도는 1이다. 신뢰관계가 있는 경우 사용자 영향도는 1보다 큰 값을 갖게 된다.

이웃사용자 유사도 조정단계에서는 사용자 영향

도를 반영하여 추천 대상자 x 의 이웃사용자 y 의 유사도를 이웃사용자 y 의 영향도를 반영하여 식(7), 식(8)과 같이 조정한다. α 는 0보다 큰 값으로 데이터의 특성에 맞게 최적의 값을 실험을 통해 찾게 된다.

$$s^*(x,y) = \alpha \times imp(y) \times s(x,y) \quad (7)$$

$$s^*(x,y) = imp(y)^\alpha \times s(x,y) \quad (8)$$

본 연구에서는 식(7)과 식(8)을 이용하여 유사도를 조정하고 실험을 통해 예측 정확도가 향상되는지를 살펴보고자 한다.

선호도 예측단계에서는 식(7), 식(8)과 식(2)를 이용하여 추천 대상자 x 를 위한 추천 대상 상품 i 의 평가점수를 예측한다.

$$p(x,i) = \bar{v}_x + \frac{\sum_{k \in N} s^*(x,k) \times (v_{ki} - \bar{v}_k)}{\sum_{k \in N} |s^*(x,k)|} \quad (9)$$

추천 대상자를 위한 상품별 선호도 예측작업이 끝나면 선호도가 높은 상품순으로 추천 대상자에게 상품을 추천하게 된다.

4. 실험분석

4.1 실험데이터

본 연구에서 제안된 알고리즘을 평가하기 위해 LibRec[17]에서 제공하는 filmtrust 데이터를 사용하였다[9]. filmtrust 데이터셋은 사용자가 영화에 대해 0.5점에서 4점사이의 영화평점을 매긴 ratings.txt와 일부 사용자가 다른 사용자에게 신뢰를 나타낸 경우 이 신뢰관계를 정리한 trust.txt 파

일로 구성되어 있다. ratings.txt는 <userid, movieid, movieRating> 3개의 속성과 35,497개의 사례로 구성되어 있고 데이터 밀도는 1.14%이다. trust.txt는 <trustor, trustee, trustRating> 3개의 속성과 1,853 사례로 구성되어 있다. 신뢰관계가 있는 경우 trustRating은 1의 값을 갖는다.

4.2 실험 및 결과분석

본 연구에서 제안한 알고리즘의 예측 정확도 평가를 위한 평가척도로 협업필터링 연구에서 일반적으로 많이 이용하는 MAE(Mean Absolute Error)를 사용하였다[1,6]. 실험데이터를 훈련데이터와 검증데이터로 구분하고 훈련데이터를 이용하여 검증데이터내의 실제 평가점수를 예측하고 MAE를 측정하였다. 평가를 위한 실험은 R언어 기반의 RStudio 환경하에서 실시하였다. 사용자간 신뢰관계는 pajek을 사용하여 사회연결망을 구성하고 구조적 공백기법을 적용하였다.

실험 알고리즘은 다음 3가지를 대상으로 하였다. 첫 번째는 전통적인 사용자기반 협업필터링(CF) 방식이고, 두 번째는 식(7)에 의해 조정된 유사도를 사용하는 방식(LinearImpCF)이며, 세 번째는 식(8)에 의해 조정된 유사도를 사용하는 방식(ExpImpCF)의 예측 정확도이다.

먼저 CF 방식의 적용에 앞서 최적 이웃사용자 수 N 을 결정하는 것이 필요하다. 추천 대상자를 위한 영화 i 의 평가점수를 계산하기 위해 추천 대상자와 유사도가 높은 순서로 영화 i 에 평점을 부여한 이웃사용자 N 의 크기를 50에서부터 700까지 50단위로 변화시켜가면서 MAE의 변화를 살펴보았다. ratings.txt의 일부 정리된 데이터에 대해 R언어의 cvTools 패키지를 활용하여 10-겹 상호검증을 실시하였다. 이웃사용자 수 N 에 따른 평균 MAE는 <그림 2>와 같다.

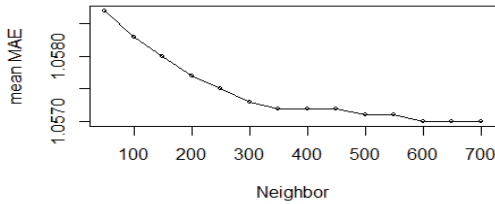


그림 2. 이웃사용자와 평균 MAE
Figure 2. Neighbor and mean MAE

10-겹 상호검증의 결과 이웃사용자의 수가 600 부터 평균 MAE가 최저를 나타냈고, 향후 실험에서는 이웃사용자 수 N 을 600으로 설정한다. ratings.txt 데이터에 대한 CF 방식의 10-겹 상호검증의 결과 MAE 는 다음 <표 1>과 같이 나타났다.

표 1. CF방식의 10-겹 상호검증 결과
Table 1. Results of 10-fold cross validation of CF

fold	MAE
1	0.9778
2	1.0287
3	1.0080
4	1.0217
5	1.0370
6	1.0079
7	1.0193
8	0.9664
9	1.0300
10	1.0925
mean	1.0189

식(7)의 유사도 조정을 사용하는 LinearImpCF 방식에서 사용할 α 를 결정하기 위해 이웃사용자 수는 600으로 하여 10-겹 상호검증을 실시하였다. 신뢰기반 사회연결망은 trust.txt 데이터를 이용하여 방향 네트워크를 구성하였고, 구조적 공백기법을 사용하여 각 노드(사용자)의 제약값을 구한후 식(6)을 이용하여 각 노드(사용자)의 영향도를 구하였다. α 를 0.5부터 4까지 변화시키며 실험한 결과, α 의

변화에 따른 정확도의 차이는 나타나지 않았다. <표 2>는 $\alpha = 1$ 인 경우의 10-겹 상호검증의 결과이다.

표 2. LinearImpCF방식의 10-겹 상호검증 결과
Table 2. Results of 10-fold cross validation of LinearImpCF

fold	MAE
1	0.9783
2	1.0275
3	1.0122
4	1.0170
5	1.0314
6	0.9942
7	1.0332
8	0.957
9	1.0270
10	1.0893
mean	1.0167

표 3. ExplmpCF방식의 10-겹 상호검증 결과
Table 3. Results of 10-fold cross validation of ExplmpCF

fold	MAE
1	0.9768
2	1.0258
3	1.0077
4	1.0168
5	1.0319
6	0.9978
7	1.0259
8	0.959
9	1.0269
10	1.0875
mean	1.0157

식(8)의 유사도 조정방식의 ExplmpCF 방식에서 사용하는 α 를 결정하기 위해 α 를 0.1~8까지 변화시키며 10-겹 상호검증을 실시하였고 그 결과 α 가 0.6인 경우 최적의 결과를 보였다. ExplmpCF방식에서 $\alpha = 0.6$ 으로 설정하고 10-겹 상호검증한 결과는 <표 3>과 같다.

10-겹 상호검증에서는 전체 데이터셋을 랜덤하게 10그룹으로 나누고 각 그룹이 한번씩 검증데이

터로 사용된다. <표 1>에서 <표 3>의 결과를 보면 검증데이터에 따라 MAE가 약간씩 차이가 있으나 전반적으로 ExplmpCF 방식의 MAE가 가장 좋을 수 있다. 실험결과 전통적인 협업필터링 방식에 비해 사용자간 신뢰관계를 반영한 협업필터링 방식의 예측 정확도가 개선되었음을 알 수 있다. 정확도의 개선이 통계적으로 유의성을 갖는 지를 확인하기 위해 R언어의 cvTools 패키지를 사용하여 10-겹 상호검증을 3회씩 실시하였고, 쌍체 t검정을 실시하여 <표 4>의 결과를 얻었다.

표 4. 쌍체 t-검정 결과
Table 4. Results of paired t-test

		mean	var.	t-value	p-value
1	CF	1.0125	0.00083	3.1111	0.0042
	LinearImpCF	1.0082	0.00084		
2	CF	1.0125	0.00083	5.2626	0.00001
	ExplmpCF	1.0079	0.00082		
3	LinearImpCF	1.0082	0.00084	0.5543	0.5837
	ExplmpCF	1.0079	0.00082		

쌍체 t-검정의 결과 LinearImpCF와 ExplmpCF 방식은 전통적 협업필터링(CF) 방식에 비해 예측 정확도를 개선하고 있고 통계적으로 유의한 것으로 나타났다. ExplmpCF 방식은 LinearImpCF에 비해 평균 MAE를 약간 개선시키나 통계적으로는 유의하지 않았다.

5. 결 론

추천 알고리즘의 예측 정확도를 높이는 것은 상품추천시스템 영역에서 주요 과제 중 하나이다. 상품추천시스템에서 많이 활용되고 있는 방법으로 협업필터링 방식이 있다. 최근들어 사회연결망 분석기법을 활용하여 추천 정확도를 높이기 위한 연구가 이루어지고 있는 데, 분석 기법으로 중심성을 활용한 연구가 주로 이루어지고 있다. 본 연구에

서는 협업필터링 방식의 예측 정확도를 높이기 위해서 신뢰관계 기반 사회연결망을 생성하고 구조적 공백기법을 적용하는 방법에 대해 제안하였다. 일반적으로 사용자는 신뢰관계가 있는 사용자의 추천을 선호하는 것으로 알려져 있다[6]. 구조적 공백은 연결망에서 유리한 위치를 점하고 있는 노드(사용자)를 찾아내며, 유리한 위치를 가진 사용자는 연결망에서 사용자의 영향도가 높게 나타난다. 본 연구에서는 추천 대상자의 이웃사용자 유사도에 이웃사용자의 사회연결망에서의 영향도를 반영하여 조정하였다. 조정된 유사도는 신뢰관계를 표현하지 않은 추천 대상자에게도 적용되는 데, 이는 명시적으로 신뢰관계를 표현하지 않았지만 암묵적으로 연결망에서 많은 신뢰를 받고 있는 사용자를 신뢰한다는 의미를 담고 있다고 볼 수 있다. 일반적인 조직에서 구조적 공백의 위치를 점하는 사람이 더 좋은 성과를 나타내는 것으로 알려져 있는데[8], 본 연구결과를 보면 신뢰관계 사회연결망에서도 구조적 공백의 위치에 있는 사람의 정보를 활용하면 추천시스템의 예측 정확도를 높일 수 있음을 알 수 있다.

filmtrust 데이터셋에 제안된 알고리즘을 적용하여 실험한 결과 추천 정확도가 개선됨을 보였다. 본 연구에서는 사전연구[9]에서 정리된 사용자간의 신뢰관계 데이터를 활용하였다. 실제 상황에서는 신뢰관계 데이터를 수집하여야 하는 선행작업이 필요하다. 인터넷 환경하에서 축적된 사용자 정보를 활용하여 사용자간 신뢰관계를 유추할 수 있는 연구가 향후 이어진다면 본 연구는 좀 더 일반적인 상황에 확장성을 가질 수 있을 것으로 보인다.

References

[1] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, *Item-based collaborative filtering*

- recommendation algorithms*, Proceedings of the 10th International Conference on World Wide Web, pp. 285-295, 2001.
- [2] K. Goldberg, T. Roeder, D. Gupta, and C. Perkins, *EigenTaste: A constant time collaborative filtering algorithm*, Information Retrieval, Vol. 4, No. 2, pp. 133-151, 2001.
- [3] H. S. Park, I. K. Eom, and Y. I. Kim, *Design and implementation of a responsive website for personalized news article recommendation and newspaper production system*, Journal of Knowledge Information Technology and Systems, Vol. 10, No. 2, pp. 205-214, 2016.
- [4] J. Goldbeck, *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, pp. 93-104, 2006.
- [5] M. G. Kim, and K. J. Kim, *Recommender system using structural hole and collaborative filtering*, Journal of Intelligence and Information System, Vol. 20, No. 4, pp. 107-120, 2014.
- [6] S. Choi, K. Y. Kwahk, and H. Ahn, *Enhancing predictive accuracy of collaborative filtering algorithms using the network analysis of trust relationship among users*, Journal of Intelligence and Information System, Vol. 22, No. 3, pp. 113-127, 2016.
- [7] P. Chen, S. Dhanasobhon, and M. D. Smith, *An analysis of the differential impact of reviews and reviewers at amazon.com*, Proceedings of 27th International Conference on Information Systems, pp. 1-17, 2007.
- [8] Y. H. Kim, *Social network analysis*, Parkyoungsa, Seoul, 2007.
- [9] G. Guo, J. Zhang, and N. Yorke-Smith, *A novel bayesian similarity measure for recommender systems*, Proceedings of the 23th International Joint Conference on Artificial Intelligence, pp. 2619-2625, 2013.
- [10] J. Herlocker, J. Konstan, A. Borchers, and J. Riedl, *An algorithm framework for performing collaborative filtering*, Proceedings of the 1999 Conference on Research and Development in Information Retrieval, pp. 230-237, 1999.
- [11] D. W. Son, *Social network analysis*, Kyungmoonsa, Seoul, 2002.
- [12] J. H. Park, 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, pp. 123-140, 2009.
- [13] Y. H. Cho, and J. H. Bang, *Social network analysis for new product recommendation*, Journal of Intelligence and Information Systems, Vol. 15, No. 4, pp. 177-193, 2009.
- [14] B. S. Kang, *A novel web recommendation method for new customers using structural holes in social networks*, Journal of Industrial Economics and Business, Vol. 23, NO. 5, pp. 2371-2385, 2010.
- [15] B. S. Kang, *Structural holes method in social networks for new product recommendations*, Journal of the Korean Data Analysis Society, Vol. 13, No. 3, pp. 1365-1377, 2011.
- [16] R. S. Burt, *Structural holes: The social structure of competition*, Harvard University Press, 1992.
- [17] LibRec, <http://www.librec.net/datasets.html>. Oct. 2016.

신뢰관계 기반 사회연결망의 구조적 공백을 활용한 추천 알고리즘의 예측 정확도 개선

강부식

목원대학교 서비스경영학부

요 약

추천 알고리즘의 예측 정확도를 높이는 것은 상품 추천시스템 영역에서 주요 과제 중 하나이다. 상품추천시스템에서 많이 활용되고 있는 방법에 협업필터링이 있다. 사용자 기반 협업필터링은 선호도 유사성이 높은 이웃사용자의 상품 선호 정보를 이용하여 상품을 추천한다. 신뢰관계나 친구관계가 있는 사용자간에는 이런 관계 정보를 추천에 활용할 경우 추천시스템의 예측 정확도가 개선될 수 있음을 보이는 연구가 최근 발표되고 있다. 본 연구에서는 협업필터링 방식의 예측 정확도를 높이기 위해서 신뢰관계 기반 사회연결망을 생성하고 구조적 공백을 적용하는 방법에 대해 제안하였다. 신뢰관계를 나타낸 데이터를 이용하여 사회연결망을 구성하고 사회연결망 분석 기법 중 하나인 구조적 공백을 적용하여 연결망에서의 사용자영향도를 찾아낸다. 추천 대상자의 이웃사용자 유사도는 찾아진 이웃사용자의 영향도를 반영하여 조정된다. 유사도를 조정하는 방식으로 두 가지를 실험하였다. LinearImpCF는 이웃사용자의 영향도에 α 를 곱하고, 이 값을 기존의 유사도에 곱하여 조정한다. ExplmpCF는 이웃사용자의 영향도에 α 승을 하고 이 값을 기존의 유사도에 곱하여 조정한다. 제안한 알고리즘의 검증은 위해 filmtrust 데이터셋에 제안된 알고리즘을 적용하여 실험하였다. 10-겹 상호검증 결과 LinearImpCF와 ExplmpCF의 평균 MAE가 전통적인 협업필터링 방식인 CF보다 낮았다. 연구에서 제안한 LinearImpCF와 ExplmpCF의 방법이 예측 정확도를 개선했음을 알 수 있다. 통계적 유의성을 검정하기 위해 10-겹 상호검증을 3회 실시하였다. 실험결과를 바탕으로 쌍체 t-검정을 실시한 결과 통계적으로 유의함을 확인하였다. 결론적으로 전통적인 협업필터링 방식에 사용자 간의 신뢰관계를 사회연결망으로 구성하고, 구조적 공백을 활용하여 사용자의 영향도를 찾아 반영하는 제안된 추천 알고리즘이 예측 정확도를 개선시킬 수 있었다.



Boo Sik Kang received the bachelor's degree in the Department of Industrial Engineering from the KyungHee University in 1985. He received the M.S. degree and the Ph.D. degree in the Department of Industrial Engineering from KAIST in 1989 and 2000, respectively. From 1989 to 2001, he was a researcher at Korea Telecom. He was a professor in the Division of Service Management at MokWon University from 2001 to 2017. His current research interests include data mining, customer relationship management, service quality management. He is a life member of the KKITS.

E-mail address: bookang@mokwon.ac.kr