

# 유전자 알고리즘을 활용한 소셜네트워크 기반 하이브리드 협업필터링\*

노희룡

국민대학교 비즈니스IT전문대학원  
(nohr@kookmin.ac.kr)

최슬비

국민대학교 비즈니스IT전문대학원  
(seulbimon@kookmin.ac.kr)

안현철

국민대학교 비즈니스IT전문대학원  
(hcahn@kookmin.ac.kr)

본 연구는 사용자 평점 이외에 사용자 간 직접·간접적 신뢰 및 불신 관계 네트워크의 분석 결과를 추가로 반영한 새로운 하이브리드 협업필터링(Collaborative filtering, CF) 추천방법을 제안한다. 구체적으로 사용자 간의 유사도를 계산할 때 사용자 평가점수의 유사성만을 고려하는 기존의 CF와 다르게, 사용자 신뢰 및 불신 관계 데이터의 사회연결망분석 결과를 추가적으로 고려하여 보다 정교하게 사용자 간의 유사도를 산출하였다. 이 때, 사용자 간의 유사도를 재조정하는 접근법으로 특정 이웃 사용자가 신뢰 및 불신 관계 네트워크에서 높은 신뢰(또는 불신)를 받을 때, 추천 대상이 되는 사용자와 해당 이웃 간의 유사도를 확대(강화) 또는 축소(약화)하는 방안을 제안하고, 더 나아가 최적의 유사도 확대 또는 축소의 정도를 결정하기 위해 유전자 알고리즘(genetic algorithm, GA)을 적용하였다. 본 연구에서는 제안 알고리즘의 성능을 검증하기 위해, 특정 상품에 대한 사용자의 평가점수와 신뢰 및 불신 관계를 나타낸 실제 데이터에 추천 알고리즘을 적용하였으며 그 결과, 기존의 CF와 비교했을 때 통계적으로 유의한 수준의 예측 정확도 개선이 이루어짐을 확인할 수 있었다. 또한 신뢰 관계 정보보다는 불신 관계 정보를 반영했을 때 예측 정확도가 더 향상되는 것으로 나타났는데, 이는 사회적 관계를 추적하고 관리하는 측면에서 사용자 간의 불신 관계에 대해 좀 더 주목해야 할 필요가 있음을 시사한다.

**주제어** : 추천시스템, 협업필터링, 신뢰, 불신, 사회연결망분석, 내향 연결정도 중심성, 유전자 알고리즘

논문접수일 : 2017년 3월 14일    논문수정일 : 2017년 3월 14일    게재확정일 : 2017년 3월 20일  
원고유형 : 학술대회우수논문    교신저자 : 안현철

## 1. 서론

추천시스템(Recommender system)은 특정 상품에 대한 구매여부나 평가점수와 같은 사용자의 데이터를 바탕으로 상품 및 서비스를 선택적으로 추천함으로써 구매의사결정 과정에서 정보 과부화(information overload) 현상을 해소하는 것

뿐만 아니라, 소비자에게 필요한 정보를 제시해주는 정보시스템을 뜻한다(Schafer et al., 2001). 추천시스템을 구현하는 방법으로는 여러 가지 기술이 있는데, 그 중에서도 협업필터링은 가장 우수한 성능을 보이며 대표적인 추천 알고리즘으로 평가받고 있다. 협업필터링 알고리즘을 기반으로 추천 서비스를 제공하고 있는 사례로는

\* 본고는 지난 2016년 한국지능정보시스템학회 추계학술대회에서 최우수논문상을 수상하여 Fast Track으로 추천되었던 학술대회 발표 논문을 확장, 보완한 논문입니다.

해의 시장의 경우 Netflix, CDNow.com, Amazon.com 등이 있으며, 국내 시장의 경우 영화 추천 애플리케이션 ‘왓챠(Watcha)’와 인터넷 중고서점 ‘알라딘’ 등이 대표적으로 알려져 있다(Son et al., 2015).

협업필터링은 내용기반 필터링이나 지식기반 필터링과 같은 다른 추천 기술들과 비교했을 때, 특정 상품의 특징을 고려하지 않아도 상대적으로 높은 정확도를 보이는 장점이 있지만, 데이터 희박성(sparsity), 확장성(scalability)의 문제 등과 같은 한계점을 갖는다. 이에 협업필터링 기반 추천시스템의 한계점을 극복하기 위해 새로운 정보의 분석 결과를 추천 알고리즘에 결합하는 하이브리드 접근법에 대한 연구들이 다양하게 시도되고 있다(Ahn, 2014).

이처럼 새로운 정보를 추가로 반영하여 추천시스템의 예측 정확도를 개선시키려는 하이브리드 협업필터링 연구 중에서도 최근 발표된 Choi et al.(2016)과 Noh and Ahn(2017)은 사용자 간의 신뢰관계 네트워크 정보를 추가로 활용하여 협업 필터링의 예측정확도 제고를 꾀한 연구들이다. 이 중, Choi et al.(2016)은 사용자 간의 신뢰관계 네트워크 분석의 결과를 사용자 간 유사도 산출 과정에 반영하여 전체적인 유사도를 재조정함으로써 전통적인 협업필터링과 비교했을 때 예측 정확도가 개선됨을 제시했는데, 이 연구는 사용자 간 신뢰 관계에만 한정하여 접근하였다는 한계가 있었다. 이후 발표된 Noh and Ahn(2017)은 Choi et al.(2016)의 연구를 확장하여, 사용자 간의 신뢰 관계 뿐만 아니라, 불신 관계도 함께 고려하는 새로운 하이브리드 추천시스템을 제안하였다.

본 연구 역시 이러한 Noh and Ahn(2017)의 연구와 마찬가지로 사용자의 신뢰와 불신 관계를

동시에 고려하도록 설계되었다. 하지만, 직접 신뢰 및 불신 관계만 사회연결망분석의 대상으로 고려하였던 Noh and Ahn(2017)의 연구와 달리, 본 연구에서는 간접 신뢰 및 불신 관계까지 사회연결망분석의 대상으로 고려하고 있다는 점에서 차별화된다.

구체적으로 본 연구가 제안하는 추천 알고리즘은 사용자 간의 유사도를 산출할 때 직접 신뢰(불신)관계, 간접 신뢰(불신)관계 이렇게 총 4개의 관계 네트워크에 대한 사회연결망분석 결과를 추가적으로 반영한다. 이 때, 새롭게 반영된 정보에 의해 사용자 간의 유사도가 재조정되는 과정에서, 유사도를 확대하거나 축소하는 4개의 조정승수의 최적값을 찾는 방법으로 유전자 알고리즘을 적용하였다. 이 때, 신뢰 및 불신 관계 데이터를 분석하기 위해서 사회연결망분석 기술 중 하나인 내향 연결정도 중심성 척도를 계산하였다.

본 논문은 다음과 같은 순서로 전개된다. 우선 2장에서는 추천시스템, 사회연결망분석, 그리고 유전자 알고리즘과 관련하여 본 연구의 토대를 이루는 기초 개념들에 대해 소개하고, 추천시스템의 성능 개선을 시도한 다양한 선행연구를 살펴본다. 3장에서는 유전자 알고리즘을 활용한 사용자 신뢰 및 불신 관계 네트워크 기반 협업필터링 연구모형을 제안한다. 4장에서는 본 연구의 실증분석을 위한 데이터, 실험 설계, 그리고 실험 결과에 대하여 논의한다. 마지막으로 5장에서는 본 연구의 결론과 함께 후속적인 연구를 통하여 극복해야 할 한계점과 향후 연구방향에 대하여 제시될 것이다.

## 2. 이론적 배경

### 2.1 추천시스템

90년대 말 이후, 인터넷의 급격한 확산은 소비자들에게 많은 양의 정보를 제공하며, 좀 더 나은 상품을 구매할 수 있도록 도왔다. 하지만 너무 과다하게 제공되는 정보의 양은 오히려 소비자들이 올바른 구매의사결정을 내리는 것을 더 어렵게 만드는 문제로 이어졌다. 이러한 부작용을 효과적으로 해결할 수 있도록, 추천시스템은 소비자가 진정으로 필요로 하는 특정 상품이나 서비스에 대한 추천결과를 생성한다(Lee and Park, 2007). 오늘날 학계는 물론 산업계에서도 많은 관심을 보이고 있는 추천시스템은, 주요한 지능형 의사결정지원시스템 분야의 주제 중 하나이다.

추천시스템은 추천결과를 생성하는 기법(알고리즘)에 따라 다양하게 구분된다(Son et al., 2015). 구체적으로 추천시스템을 구현하는 방법으로는 베스트셀러 기반 기법, 내용 기반(content-based, CB) 기법, 인구통계정보 기반 기법, 최소질의대상 기법, 협업필터링(collaborative filtering, CF) 등이 있는데, 이 중에서도 협업필터링이 산업계나 학계에서 가장 많이 활용 및 연구되고 있으며, 대표적인 추천결과 생성 알고리즘으로 평가받고 있다.

협업필터링은 특정 상품에 대한 구매 이력 혹은 사용자가 입력한 평점 데이터를 가지고 사용자가 선호할 것으로 예상되는 적합한 상품을 추천해 주는 기법(Adomavicious and Tuzhilin, 2005; Cho and Kim, 2004; Sarwar et al., 2000)으로, 미네소타 대학의 GroupLens 프로젝트를 통해 처음으로 소개된 이후, 오늘날 MovieFinder,

Amazon.com과 CDNow.com 그리고 인터넷 중고서점인 ‘알라딘’, 영화 추천 애플리케이션 ‘왓차’ 등 국내·외 다수의 전자상거래 업체들에 의해 채택되어 사용되고 있다.

전통적인 협업필터링 기반 추천 프로세스는 다음의 <Table 1>에 제시된 것과 같이 기본적으로 사용자-상품 평점 행렬을 구성하고, 이를 기반으로 추천 대상에게 추천할 적합한 상품을 결정한다.

<Table 1> User-Item Rating Matrix

	A Item	B Item	C Item	D Item	E Item
User 1	5		2		2
User 2	4	2		3	5
User 3	3	3	3		1
User 4	5	3		2	3
User 5	4	2		3	??

예를 들어, <Table 1>에서 협업필터링을 적용해 'User 5'의 'E Item'에 대한 예상 선호도를 추정한다고 하자. 이 경우, 'User 5'는 해당 사용자를 제외한 나머지 4명의 사용자 중 'User 2'와 'A Item' ~ 'D Item'에 대한 평점 패턴이 완전히 동일한 것으로 나타나고 있는데, 이는 동일한 평점 패턴을 가진 'User 5'와 'User 2'가 유사한 선호체계를 가졌다고 볼 수 있다. 따라서 'User 2'가 'E Item'에 대해 5점을 부여하였기 때문에, 'User 5' 역시 'E Item'에 5점이라는 높은 평점을 부여할 것이라고 예측할 수 있다. 이처럼 협업필터링은 상품의 특징(내용)을 이해할 필요 없이, 단순히 사용자들의 평점 패턴 또는 구매여부와 같은 구매이력만으로 쉽게 추천리스트를 생성할 수 있다는 것이 가장 큰 장점이라고 할 수 있다.

이처럼 협업필터링은 장점이 많아 활용도가 높지만, 한편으로 데이터 희소성(sparsity), 확장성(scalability) 등의 문제점이 협업필터링의 한계점으로 지적되고 있다. 이에 다양한 기법들과의 결합을 고려하는 하이브리드 접근법을 통해 이러한 협업필터링의 한계점을 극복하려는 시도가 오래 전부터 존재해 왔다(Burke, 2002; 2007; McNee et al., 2006). 아울러 최근에는 하이브리드 접근법을 통해 기존 협업필터링의 한계점 극복 뿐 아니라 관점을 더욱 확장시켜 기존 협업필터링의 정확도를 한층 제고하기 위한 시도들도 활발하게 이루어지고 있다(Choi et al., 2016; Ganu et al., 2009; Jeon and Ahn, 2015; Ling et al., 2014; McAuley and Leskovec, 2013; Vozalis and Margaritis, 2004).

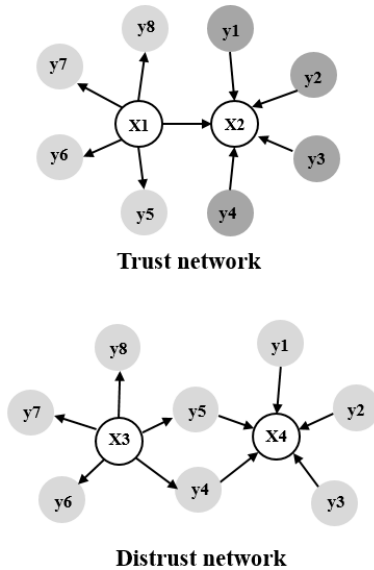
## 2.2 사회연결망분석

Barnes(1954)에 의해 처음으로 제안된 사회연결망분석(social network analysis, 이하 SNA)은 네트워크 구조에 존재하는 노드(node) 각각의 개별적인 특성보다는, 네트워크의 전체적인 연결 구조나 연결 상태에 초점을 맞추고 그 패턴을 분석하는 기법이다(Ryu et al., 2006). 이러한 SNA는 인간의 관계 네트워크를 중심으로 인간의 행동 범위와 사회적 구조를 설명하려는 시도에서 출발되어(Kim, 2003; Shin et al., 2012), 그 활용 범위가 점차 확대되어 오늘날 의학, 마케팅, 물리학 등 다양한 분야에서 많은 관심을 받고 있다. 특히 컴퓨팅 기술이 발전됨에 따라 대규모 데이터의 분석이 가능해 지면서, SNA를 다양한 분야에 적용한 연구들이 활발하게 발표되고 있다(Ahn et al., 2007; Mislove et al., 2007; Thelwall, 2008; Wang et al., 2009).

SNA에는 네트워크 구조를 파악할 수 있는 다양한 척도가 존재하는데, 그 중 중심성(closeness)은 전체 네트워크 구조에서 어떤 노드가 중심에 위치하는지에 대한 정도를 의미한다(Choi et al., 2016). 중심성에는 크게 연결정도(degree) 중심성, 근접(closeness) 중심성, 매개(betweenness) 중심성, 아이겐벡터(Eigenvector) 중심성 등이 있다(Kwahk, 2014).

본 연구에서는 위에서 소개한 다양한 중심성 척도 중, 가장 간단하게 중심성을 측정할 수 있는 방법으로 알려져 있는 연결정도 중심성 척도를 기반으로 연구를 진행하였다(Jeong and Kim, 2013). 연결정도 중심성은 가장 대표적인 중심성 척도로, 한 개의 노드(액터)와 직접 연결 관계를 갖고 있는 노드 개수의 합으로 구할 수 있다. 사회연결망은 방향성의 유무에 따라 비방향(nondirected) 네트워크와 방향(directed) 네트워크로 나뉠 수 있는데, 연결정도 중심성 척도는 비방향 네트워크의 경우 단순히 연결 관계 개수의 합으로 중심성이 평가되고, 방향 네트워크의 경우에는 내향·외향 연결정도로 구분되어 중심성이 평가된다(Choi et al., 2016). 아래 <Figure 1>에 제시된 바와 같이, 본 연구의 대상이 되고 있는 신뢰 및 불신 관계 네트워크는 방향성이 존재하는 방향 네트워크의 일종이다. 이러한 방향 네트워크에서 외향 연결정도(outdegree)는 해당 노드(액터)가 얼마나 개방적인 성향(expansiveness)을 갖고 있는지를 의미하고, 내향 연결정도(indegree)는 해당 노드(액터)의 인기 정도를 나타낸다(Kwahk, 2014).

초기의 연구들을 살펴보면, 사회연결망분석을 본격적으로 사용하기보다는 단순히 사회연결망 정보를 탐색하고 이를 추천과정에 반영하려는 시도에서 그쳤다(Noh and Ahn, 2017). 사회연결



(Figure 1) Examples of outdegree and indegree trust/distrust network

망 정보를 활용해 FilmTrust라는 이름의 새로운 영화 추천시스템을 제안한 Golbeck(2006)은 평균과 비교했을 때, 사용자들의 평점 데이터가 다양하게 존재할 때 추천의 정확도가 높아진다고 주장하였다.

최근에는 초기의 연구들과 다르게 네트워크 탐색 정보를 반영하는 것이 아닌, 사회연결망분석을 활용해 산출된 결과물을 추천 과정에 반영한다. 예를 들어, Cho and Bang(2009)은 사용자 구매정보 네트워크를 구축한 후, 중심성 척도를 계산한 결과를 바탕으로 새로운 상품을 추천하는 알고리즘을 제안하였으며, Shin et al.(2012)은 중심성 분석기법을 결합한 하이브리드 접근법을 제시하여 신규 사용자에 대한 상품 추천시스템의 예측 정확도를 제고하고자 하였다.

### 2.3 신뢰 정보를 결합한 추천시스템

신뢰는 연구자들 사이에서 접근 관점에 따라 다양하게 해석되고 있는데, 여기서 신뢰는 주어진 상황에서 신뢰하고자 하는 대상의 행동을 과연 믿을 수 있는지에 대한 의지로 해석되며, 신뢰대상이 권위에 따른 기회주의적 행동을 하지 않을 것이라는 믿음을 근간으로 한다(Hahn and Kim, 2009; Hurley, 2011; Mayer et al., 1995). 신뢰는 관계 구조에 있어서 서로에 대한 신뢰를 가질 수 있기 때문에 네트워크 관계에 있어서 방향이 존재한다. Lu et al.(2011)은 어떤 대상에 대한 신뢰가 그와 다른 대상을 인식할 때의 관점과 태도에 영향을 미친다고 하였는데, 이러한 인지 과정을 신뢰의 전이로 정의하였다.

Du et al.(2016)은 신뢰대상 뿐만 아니라 그 대상이 신뢰하는 새로운 대상들을 간접적으로 신뢰하는 것에 있어서 신뢰의 전이가 큰 영향을 미친다고 주장하였으며, Yuan(2010)은 전자상거래의 급격한 성장에 따라 거래 위험과 거래 불확실성에 대한 우려가 점차 커지는 가운데, 신뢰의 전이가 이러한 문제들을 해소시켜주는 중요한 요인으로 작용하고 있다고 주장하였다.

신뢰 인식 추천시스템(trust-aware recommender system, 이하 TARS)은 신뢰와 신뢰전이가 구매 의사결정에서 직접 또는 간접적인 영향을 줄 수 있다는 점을 활용해, 사용자의 구매를 돕는 가치 있는 정보를 제공하는 추천시스템을 말한다. 이를 기반으로 한 선행연구들을 살펴보면 다음과 같다.

Massa and Avesani(2004)는 데이터 희소성 문제해결에 초점을 맞추고, 기존의 TARS 보다는 적은 비용으로 내재적 신뢰를 산출하고 전파함과 동시에 더 우수한 평가 예측 정확도와 더 합

리적인 평가 예측 범위를 갖는 추천 시스템을 제안하였다. Wei(2013)는 직접적인 관련이 없는 새로운 대상에 대한 신뢰를 추론하는 것이 TARS 연구의 핵심 주제라고 주장했고, Yuan(2010)은 신뢰의 추론을 통해 TARS의 평가 예측 범위와 정확도를 개선할 수 있다고 주장하였다. Liu and Lee(2010)와 Ahn et al.(2012)는 SNS상에서 직접 또는 간접적으로 연결된 친구들의 정보를 참조한 새로운 추천 알고리즘을 제안하였다. 앞서 소개한 연구는 신뢰 관계를 직접적으로 참조한 것은 아니지만, SNS 상에서의 친구 관계는 서로에 대한 신뢰를 동반할 것이라고 보았다. 그 결과, 제안한 추천 알고리즘의 예측 정확도가 개선됨을 확인할 수 있었다. Ha(2014)가 제안한 추천 알고리즘은, 사용자들이 직접적으로 명시한 신뢰 정보와 관계 정보(예를 들어, SNS 상의 친구 관계 등)를 고려했으며, 이들의 연결강도를 모델에 반영하였다. 실험 결과, 제안 모델 중 신뢰 모델은 선호도 예측에서, 그리고 소셜 모델은 추천 상품 예측에서 우수한 예측 정확도를 보였음을 확인할 수 있었다.

최근에 발표된 Choi et al.(2016), Du et al.(2016) 그리고 Tang et al.(2016)은 신뢰 관계 정보를 직접적으로 고려했을 때 추천시스템의 예측 정확도가 향상됨을 확인할 수 있었다. 특히, Choi et al.(2016)의 연구는 신뢰 관계 네트워크 정보의 내향 및 외향 연결정도 중심성의 계산 결과를 추가로 반영하였다. 그러나 지금까지의 선행연구를 비추어 볼 때, 추천 결과를 산출함에 있어서 단순히 신뢰 관계만을 고려한 것으로 살펴볼 수 있다. 이는 본 연구가 새로운 연구방향을 제시하는 데 있어서 특별한 의미를 갖게 해준다.

본 연구는 기존의 선행연구들과는 다르게, 사

용자 간의 신뢰 관계뿐만 아니라, 오늘날 많은 연구자들 사이에서 활발하게 연구되고 있는 불신 관계도 함께 고려한 새로운 하이브리드 협업 필터링을 제안한다. 그리고 과연 신뢰 및 불신 관계 정보 중 어느 것이 더 중요한 의미를 갖는지도 살펴보고자 한다.

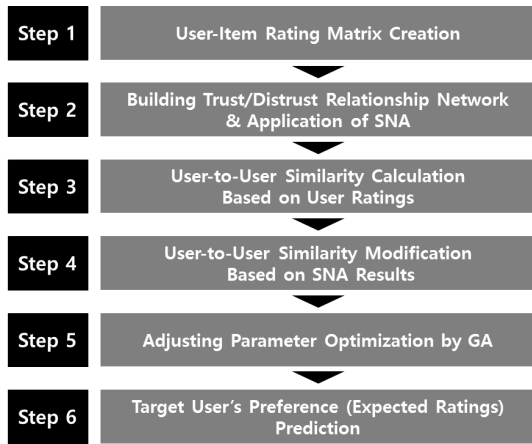
## 2.4 유전자 알고리즘

유전자 알고리즘(genetic algorithm)은 적자생존(survival of the fittest)의 원리와 유전법칙을 응용한 최적화 기법으로, 선택, 교배, 돌연변이 등으로 이루어지는 생물의 진화과정을 토대로 탐색 공간을 탐색하여 최적 또는 유사 최적해를 찾아내는 탐색기법이다(Hong and Shin, 2003). 유전자 알고리즘은 성능이 우수하기 때문에 변수와 제약이 많고 다수의 조건을 갖는 복잡한 문제를 해결할 때 유용하다. 유전자 알고리즘은 다양한 분야에서 광범위하게 활용되고 있는데, 최근에는 임계치 최적화를 위한 최적의 해를 찾는 데 널리 적용되고 있다(Ahn et al., 2009).

## 3. 제안 알고리즘

본 연구에서는 직접·간접적 신뢰 및 불신 관계 정보의 사회연결망분석 결과를 반영하여, 전통적인 협업필터링보다 더 우수한 예측 정확도를 보이는 하이브리드 추천 알고리즘을 제안한다.

본 연구에서 제안하는 추천시스템 알고리즘은 편의상 SNACF-GA(Social Network Analysis-based CF using GA)로 명명하였다. SNACF-GA는 다음의 <Figure 2>에 제시된 것처럼, 총 6단계의 절차에 따라 구현된다.



(Figure 2) Procedure of SNACF-GA

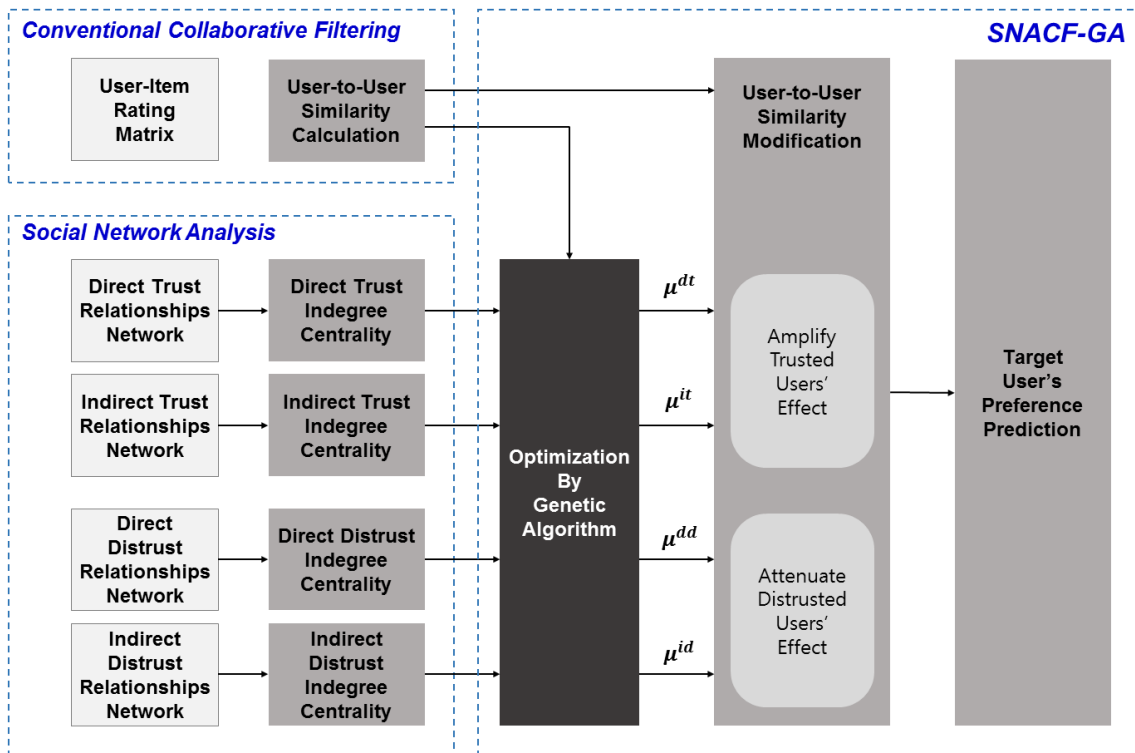
<Figure 3>는 전반적인 SNACF-GA의 구조체

계를 나타내고 있다. 이 그림에서 볼 수 있듯이, SNACF-GA는 전통적인 CF와 SNA, 그리고 유전자 알고리즘에 기반한 최적화가 결합되어 표적 고객의 선호도를 보다 정확하게 예측할 수 있도록 설계되어 있다.

제안 알고리즘인 SNACF-GA의 작동 원리를 6 단계로 구성된 구현절차에 따라 보다 상세히 살펴보면 다음과 같다.

1단계 : 사용자-상품 평점 행렬 도출

본 연구가 제안하는 SNACF-GA의 첫 번째 단계는 사용자들의 특정 상품에 대한 평점 데이터들을 가지고, 앞서 <Table 1>에서 예시된 형태의 행렬을 도출한다.



(Figure 3) Architecture of SNACF-GA

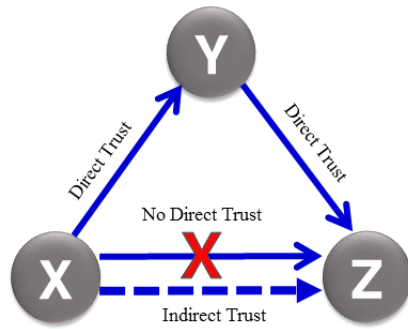
2단계 : 신뢰 및 불신 관계 네트워크 구축 및 사회연결망분석 수행

이 단계에서는 먼저 신뢰 및 불신 관계로 구성된 각각의 관계 네트워크를 구축한다. 그런 다음, 특정 사용자가 다른 사용자들에게 신뢰와 불신을 얼마나 받고 있는지를 반영하기 위해, 사회연결망분석에서 액터의 인기를 의미하는 지표인 내향 연결정도 중심성을 계산한다.

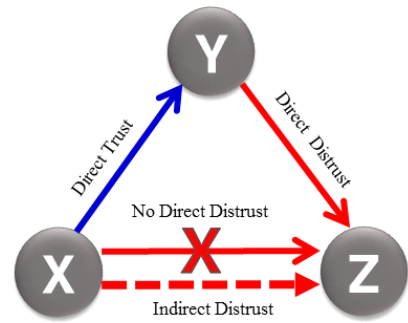
아울러 SNACF-GA에서는 신뢰 및 불신 관계 네트워크 구축 시, 직접·간접 관계를 모두 고려하였다. 직접 신뢰와 직접 불신 관계는 특정 사용자가 다른 사용자에 대해 명시적으로 해당 사용자를 신뢰 또는 불신한다고 의사를 표명한 경우를 말한다. 반면, 간접 신뢰와 간접 불신 관계는 직접적으로 의사를 표명하지는 않았지만, 특정 사용자가 신뢰하는 다른 사용자가 타 사용자를 신뢰 또는 불신할 경우, 이를 간접적으로 신뢰 또는 불신하는 것으로 간주하였다. 다음의 <Figure 4>는 이러한 4종의 관계(직접 신뢰, 직접 불신, 간접 신뢰, 간접 불신)를 잘 설명해주고 있다.

<Figure 4> 상단의 그림을 보면, 사용자 ‘X’는 사용자 ‘Y’를 직접적으로 신뢰하고, 사용자 ‘Y’는 사용자 ‘Z’를 직접적으로 신뢰하고 있다. 이때, 사용자 ‘X’는 사용자 ‘Z’를 직접적으로 신뢰하지 않고 있지만, 사용자 ‘X’가 직접적으로 신뢰하는 사용자 ‘Y’가 사용자 ‘Z’를 직접적으로 신뢰하고 있기 때문에, 사용자 ‘X’는 사용자 ‘Z’를 간접적으로 신뢰한다고 볼 수 있다. 하지만, 간접 불신 관계의 경우는 신뢰의 경우와 다르게 접근되어야 한다(<Figure 4> 하단의 그림 참고).

만약 신뢰의 경우와 마찬가지로, 사용자 ‘X’가 사용자 ‘Y’를 직접적으로 불신하고, 사용자 ‘Y’가 사용자 ‘Z’를 직접적으로 불신한다고 할 때,



(a) Indirect trust



(b) Indirect distrust

<Figure 4> Types of trust and distrust relationships

이를 사용자 ‘X’가 사용자 ‘Z’를 불신한다고 해석한다면 논리적 오류가 발생하기 때문이다. 이 경우, 사용자 ‘X’가 사용자 ‘Y’를 불신하는 상황에서 사용자 ‘Y’가 제공하는 사용자 ‘Z’에 대한 평가는 사용자 ‘X’에게 직접적으로 어떤 영향도 미치지 어렵다. 때문에, 여기서는 사용자 ‘X’가 사용자 ‘Y’를 신뢰하고, 사용자 ‘Y’가 사용자 ‘Z’를 불신할 때 (하지만, ‘X’는 직접적으로 ‘Z’를 불신하지 않을 때), 비로소 사용자 ‘X’가 사용자 ‘Z’를 간접적으로 불신한다고 해석할 수 있다. 2 단계에서는 이러한 관점을 적용하여, 직접적인 신뢰 및 불신 관계 네트워크로부터 간접적인 신



되 및 불신 관계 네트워크를 구축하게 되며, 이 네트워크에서도 내향 연결정도를 기반으로 각각 사용자의 중심성을 계산하게 된다.

### 3단계 : 사용자간 평점 유사도 산출

3단계에서는 사용자-상품 평점 행렬을 참조하여 사용자 간의 평점 유사도를 산출한다. 본 연구에서는 유사도 산출을 위해서 피어슨 상관관계 수(PCC)를 사용하였다. PCC 식은 식 (1)과 같다.

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

$S_{x,y}$ 는 사용자  $x$ 와 사용자  $y$ 의 유사도를 의미하는데, 여기서  $R_{x,i}$ 은 사용자  $x$ 의 특정 상품  $i$ 에 대한 평점을 나타내고,  $R_{y,i}$ 은 사용자  $y$ 의 특정 상품  $i$ 에 대한 평점을 나타낸다.  $\bar{R}_x$ 은 사용자  $x$ 의 평점 평균값을,  $\bar{R}_y$ 은 사용자  $y$ 의 평점 평균값을 의미한다.

### 4단계 : 사용자 간 유사도 조정

4단계에서는 2단계의 사용자 신뢰 및 불신 관계 내향 연결정도 중심성 분석 결과를 3단계에서 도출한 사용자 간 유사도와 결합하여 전체적으로 유사도를 재조정하는 작업을 거쳐, 통합적인 사용자 간 유사도를 산출한다. 새로운 사용자 간 유사도  $S^*_{x,y}$ 는 다음 식 (2)와 같이 산출된다.

$$S^*_{x,y} = \frac{amp_{x,y}^{dt} amp_{x,y}^{it}}{amp_{x,y}^{dd} amp_{x,y}^{id}} \times S_{x,y} \quad (2)$$

위 식에서  $S^*_{x,y}$ 는 신뢰 및 불신 관계 네트워

크 정보를 추가로 고려하여 사용자  $x$ 와 이웃 사용자  $y$ 의 조정된 유사도를 의미한다.  $amp_{x,y}^{dt}$ ,  $amp_{x,y}^{it}$ ,  $amp_{x,y}^{dd}$ ,  $amp_{x,y}^{id}$ 는 사용자  $x$ 와 이웃 사용자  $y$  사이의 유사도를 확대 혹은 축소시켜주는 조정계수인데, 본 연구에서는 이웃 사용자  $y$ 의 신뢰 및 불신관계 네트워크에 대한 내향 연결정도 중심성을 고려하여 유사도를 확대 혹은 축소한다. 예를 들어, 신뢰 관계 네트워크에서 사용자 'A'의 내향 연결정도 중심성 값이 높다고 가정해보자. 이는 다른 사용자들로부터 높은 신뢰를 받는 것으로 해석된다. 반대로 불신 관계 네트워크에서 사용자 'A'의 내향 연결정도 중심성 값이 높다면, 이 경우 다른 사용자들이 해당 사용자의 평가에 불신을 갖고 있다는 의미로 해석되므로, 구매의사결정 과정에서 사용자 'A'의 평점 패턴은 덜 고려하는 것이 유리하다. 본 연구에서는 신뢰 및 불신 관계를 동시에 고려해야하기 때문에, 식 (2)에 같이 총 4개(직접 신뢰, 직접 불신, 간접 신뢰, 간접 불신)의 조정계수가 필요하게 된다. 각 조정계수에 대한 산출 식은 다음의 식 (3)~(6)과 같다(Choi et al., 2016).

$$amp_{x,y}^{dt} = (1 + IC_y^{dt})^{\mu_{dt}} \quad (3)$$

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

$$amp_{x,y}^{dd} = (1 + IC_y^{dd})^{\mu_{dd}} \quad (5)$$

$$amp_{x,y}^{id} = (1 + IC_y^{id})^{\mu_{id}} \quad (6)$$

위의 식에서  $\mu$ 는 조정승수이고,  $IC_y$ 는 사용자  $y$ 의 직접 · 간접 신뢰 혹은 불신의 내향 연결정도 중심성 지수이다. 여기서 조정승수  $\mu$ 는 유사도에서 내향 연결정도 중심성을 비중 있게 반영하기 위해 결정해야하는 지표로써, 탐색을 통해 최적의 값을 찾아야 한다.

**5단계 : 유전자 알고리즘에 의한 최적 조정승수 도출**

5단계에서는 앞서 4단계에서 언급된 4개 조정승수( $\mu_{dt}, \mu_{it}, \mu_{dd}, \mu_{id}$ )의 최적값을 찾는 작업이 이루어지게 된다. 본 연구에서는 유전자 알고리즘을 이용해 승수들의 최적 값을 결정하고자 하였다. 이 단계에서는 우선 4개 조정승수에 대한 후보값들을 무작위로 생성하여, 50~100개 사이의 염색체(chromosome)들로 구성된 모집단(population)을 구성하게 된다. 그런 다음 각 염색체 별로 코딩된 승수값들을 적용하여 4단계의 유사도 조정에 기반한 협업 필터링을 수행한 뒤, 각 염색체별 성능을 측정한다. 이 때 염색체별 성능을 측정하기 위한 적합도 함수(fitness function)은 협업 필터링의 예측 정확도를 반영할 수 있는 평균 MAE(Mean Absolute Error)로 하였다. 이렇게 각 염색체별 성능이 측정되면, 그 성능에 따라 우성의 염색체를 선택(selection), 교배(crossover)하고, 간혹 돌연변이(mutation)를 일으켜 전역 최적해를 탐색하게 하는 유전 조작(genetic operation)을 수행한다. 이러한 과정을 약 50~100세대 진화시키고 나면, 결국 최적 혹은 유사최적 상태의 조정승수를 도출할 수 있게 된다.

**6단계 : 선호도 예측**

마지막 6단계에서는 5단계에서 최종 확정된 조정승수들을 기반으로 추천 대상자의 각 상품에 대한 평점(선호도)를 예측하는 작업이 이루어진다. 각 상품별 선호도( $p_{x,i}$ ) 예측은 아래 식 (7)을 사용하여 계산한다(Choi et al., 2016; Jeon and Ahn, 2015; Noh and Ahn, 2017).

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

위의 식에서  $\bar{r}_x$ 는 사용자  $x$ 의 평점의 평균값을 의미하고,  $r_{y,i}$ 는 이웃 사용자  $y$ 의 특정 상품  $i$ 에 대한 평점을 의미한다. 그리고  $S_{x,y}^*$ 는 추천 대상자  $x$ 와 이웃 사용자  $y$ 사이의 조정된 유사도로서, 앞서 식 (2)에서 도출된 유사도가 사용된다. 끝으로  $N$ 은 유사한 이웃 사용자를 탐색하는 과정에서 찾은 가장 가까운(유사한) 이웃 사용자들의 집합을 나타낸다.

식 (7)의 계산과정을 거친 후, 사용자의 선호도 예측이 완료되면, 추천 대상자에게 추천하기에 적합한 상품 리스트를 생성하고 상품을 추천한다(Choi et al., 2016; Jeon and Ahn, 2015; Noh and Ahn, 2017).

**4. 실험설계 및 실증분석**

**4.1 실험 데이터**

본 연구가 제안한 SNACF-GA의 유용성을 논리적으로 검증하기 위하여, 'trustlet.org'에서 제공하는 'Extended Epinions dataset'을 사용하였다. 이 데이터는 Epinions.com 사이트에서 사용자들이 다양한 상품(자동차, 책, 음악 등)을 구매한 후 직접 제품 리뷰를 작성하고 다른 사용자들에게 공유한 정보이다. 데이터는 지난 2001년 1월부터 2003년 8월까지 약 2년 7개월 동안 수집된 데이터로, 두 개의 파일로 구성되어있다. 첫 번째 파일명은 'ratings\_data.txt'로 사용자 ID, 상품 구매에 따른 평가 점수(5점 척도)를 부여한 내용

이 포함되어있다. 이 파일에 저장된 사용자 수는 모두 132,000명이고, 평가된 상품은 동일 상품을 제외한 총 539,460종이다. 두 번째 파일명은 'user\_rating.txt'이고, 사용자 간의 신뢰 및 불신 관계 정보를 포함하고 있다. 이 때, 사용자 간의 관계는 신뢰(1), 불신(-1)로 구분하여 나타내고 있다. 데이터는 총 841,372건이며, 이 중 약 85%에 해당되는 건수인 717,667건은 신뢰 관계를, 약 15%에 해당되는 건수인 123,705건은 불신 관계를 나타낸다.

본 연구는 SNACF-GA를 구현하기 위하여 사회연결망분석 도구 중의 하나인 UCINET 6를 사용하였으며, 사용자 간 신뢰 및 불신 네트워크의 내향 연결정도 중심성을 계산하였다. 그리고 Microsoft Excel VBA로 본 연구가 제안하는 추천 알고리즘을 구현하였다. 아울러, 앞서 언급한 4개의 조정승수의 최적값을 찾기 위한 방법으로 유전자 알고리즘을 적용하기 위해서 Excel에서 작동이 가능한 Palisade Software사의 Evolver를 사용하였다.

## 4.2 실험설계

제안 알고리즘의 성능을 확인하기 위해, 본 연구에서는 사용자의 평점이 입력된 상품들에 대해서 본 연구의 제안 알고리즘인 SNACF-GA로 예상 평점을 도출해 본 다음, 이 예상 평점과 실제 평점과의 오차를 확인하여 이 오차가 비교모형 대비 작은지를 확인해 보는 방식을 통하여 검증 진행하였다. 이 때, 평점간의 오차는 전통적으로 가장 많이 사용되어 온 MAE(Mean Absolute Error)를 사용해 측정하였다(Breese et al., 1998; Choi et al., 2016; Jeon and Ahn, 2015; Noh and Ahn, 2017; Sarwar et al., 2001).

제안 알고리즘의 성능을 보다 면밀히 파악하기 위해, 본 연구에서는 2개의 비교모형을 사용하였다. 첫 번째 비교모형은 전통적인 CF이다. 이는 오로지 평점만 고려하고, 신뢰 또는 불신 네트워크 정보는 전혀 고려하지 않는 알고리즘이다.

두 번째 비교모형은 SNACF(SNA based CF)이다. SNACF는 제안 알고리즘인 SNACF-GA와 달리, 간접 신뢰·불신 관계 네트워크를 고려하지 않는다( $\mu_{it} = \mu_{id} = 0$ 으로 설정된 것과 같음). 여기서는 오로지 직접 신뢰 및 불신 관계만을 고려하므로 조정승수도 2개( $\mu_{dt}, \mu_{dd}$ )만 사용하게 된다. 따라서 여기서는 조정승수 값 결정에 GA 최적화 대신 그리드 탐색(grid search) 기법을 사용한다.

## 4.3 실험결과

하단의 <Table 2>는 비교모형인 SNACF의 실험 결과를 정리한 결과이다. 직접 신뢰와 직접 불신만 고려하도록 설계되어 있는 SNACF에서 신뢰와 불신 관계에 대해 동시에 0의 조정승수 값을 부여(즉,  $\mu_{dt} = \mu_{dd} = 0$ ) 하게 되면, 항상  $amp_{x,y}^{dt}$ 와  $amp_{x,y}^{dd}$ 의 값이 1이 되어,  $S^*_{x,y} = S_{x,y}$ 가 된다. 즉,  $\mu_{dt} = \mu_{dd} = 0$  일 경우, SNACF는 전통적인 CF와 동일한 알고리즘이 된다.

이처럼 전통적인 CF 예측 성과를 포함하고 있는 <Table 2>의 비교모형 실험 결과를 보면, 신뢰정보를 고려한 SNACF(Avg. MAE = 0.112546)는 전통적인 CF(Avg. MAE = 0.112638) 대비 소폭의 성과개선을 보이는 것을 알 수 있다. 그리고 불신정보를 고려한 SNACF(Avg. MAE =

<Table 2> Predictive accuracy of SNACF

Type	$\mu_{dt}$	$\mu_{dd}$	Average MAE
Conventional CF	0	0	0.112638
SNACF (Trust Only)	0.2	0	0.112546
	0.4	0	0.112568
	0.6	0	0.112703
	0.8	0	0.112937
	1	0	0.113277
SNACF (Distrust Only)	0	0.2	0.112546
	0	0.4	0.112491
	0	0.6	0.112464
	0	0.8	0.112455
	0	1	0.112465
SNACF (Trust+Distrust)	0.2	0.2	0.112438
	0.2	0.4	0.112366
	0.2	0.6	0.112329
	0.2	0.8	0.112309
	0.2	1	0.112303
	0.4	0.2	0.112429
	0.4	0.4	0.112334
	0.4	0.6	0.112276
	0.4	0.8	0.112236
	0.4	1	0.112214
	0.6	0.2	0.112521
	0.6	0.4	0.112398
	0.6	0.6	0.11231
	0.6	0.8	0.11225
	0.6	1	0.112209
	0.8	0.2	0.112718
	0.8	0.4	0.112557
	0.8	0.6	0.112439
	0.8	0.8	0.112351
	0.8	1	0.112287
1	0.2	0.113007	
1	0.4	0.112811	
1	0.6	0.112661	
1	0.8	0.112545	
1	1	0.112456	

0.112455)는 신뢰정보를 고려한 SNACF보다 예측정확도가 더 우수함을 알 수 있는데, 이는 불신관계 정보가 신뢰관계 정보 보다 사용자간 유사도를 보다 정확하게 파악하는데 있어 더 가치 있는 정보임을 시사한다.

한편, 신뢰와 불신정보를 같이 고려할 경우 신뢰 중심성에 할당된 조정승수( $\mu_{dt}$ )가 0.6, 불신 중심성에 할당된 조정승수( $\mu_{dd}$ )가 1.0일 때, 가장 뛰어난 정확도(Avg. MAE = 0.112209)가 산출됨을 알 수 있다. 여기서도 최적의 불신 조정승수의 값이 신뢰 조정승수의 값보다 훨씬 높게 도출되었음을 확인할 수 있는데, 이 역시 불신관계 정보가 신뢰관계 정보에 비해 더 유의미함을 시사하는 또 다른 증거라 할 수 있다.

한편 직·간접 신뢰와 불신 관계를 모두 분석하여 추천결과를 생성하는 본 연구의 제안모형인 SNACF-GA의 실험 결과는 다음 <Table 3>에 제시되어 있다. 이 표에서 볼 수 있듯이, SNACF-GA는 직접 신뢰 0.0000와 직접 불신 1.4282의 조정계수를, 간접 신뢰 1.5000와 간접 불신 0.4615의 조정계수를 부여했을 때 가장 최적의 결과를 산출했으며, 이 때의 평균 MAE는 0.111943인 것으로 나타났다. 이는 전통적인 CF(Avg. MAE = 0.112638)에 비해 크게 개선된 예측 정확도로서, SNACF(평균 MAE = 0.112209)와 비교해도 괄목할 정도로 성과가 개선된 것임을 확인할 수 있다.

<Table 3> Prediction accuracy of SNACF-GA

Amplifying/Attenuating Coefficients				Average MAE
Direct Trust ( $\mu_{dt}$ )	Direct Distrust ( $\mu_{dd}$ )	Indirect Trust ( $\mu_{it}$ )	Indirect Distrust ( $\mu_{id}$ )	
0.0000	1.4287	1.5000	0.4615	0.111943

다음의 <Table 4>는 <Table 2>와 <Table 3>을 정리하여, 제안 알고리즘(SNACF-GA)과 2개의 비교모형(전통적인 CF, SNACF) 간 예측 정확도를 종합적으로 제시하고 있다. 이 표에 제시된 바와 같이 본 연구의 전체 실증분석 결과를 종합해 본다면, SNACF-GA > SNACF > 전통적인 CF의 순서로 예측 정확도가 높게 나타났음을 확인할 수 있다.

<Table 4> Summary of the experimental results

Algorithm	Average MAE	Setting
Conventional CF	0.112638	$\mu_{dt} = \mu_{dd} = 0$
SNACF	0.112209	$\mu_{dt} = 0.6,$ $\mu_{dd} = 1.0$
SNACF-GA	0.111943	$\mu_{dt} = 0,$ $\mu_{dd} = 1.4287,$ $\mu_{it} = 1.5,$ $\mu_{id} = 0.4615$

끝으로 알고리즘 간의 예측 정확도 차이가 과연 통계적으로 유의한 차이인지 검증하기 위하여, 대응표본 t-검정을 수행하였다. 다음의 <Table 5>는 이러한 대응표본 t-검정의 결과를 나타내고 있다. 이 표에서 알 수 있듯이, SNACF-GA 알고리즘은 전통적인 CF는 물론 SNACF와도 통계적으로 유의한 성과 차이를 보이는 것으로 나타났다. 또한, SNACF와 전통적인 CF 사이에도 통계적으로 유의한 성과차이가 있음을 확인할 수 있었다.

<Table 5> t-values of paired-samples t-test

	SNACF	SNACF-GA
Conventional CF	2.647**	2.863**
SNACF		1.983*

\* statistical significant at 5%, \*\* statistical significant at 1%

## 5. 결론

본 연구는 지금까지 신뢰 관계만을 고려했던 선행연구들과는 달리, 사용자 간 신뢰 및 불신 관계를 동시에 고려하고, 이를 분석한 결과를 추가로 활용하는 하이브리드 협업필터링인 SNACF-GA를 제안하였다. 본 연구는 제안 알고리즘의 타당성을 검증하기 위하여 웹에서 제공 받은 데이터에 적용하여 실험한 결과, 전통적인 CF는 물론 SNACF와도 비교했을 때 통계적으로 유의미한 예측 정확도의 개선을 확인할 수 있었다. 이와 같은 본 연구의 실험 결과를 바탕으로 몇 가지 시사점을 제시한다.

첫째, 본 연구는 신뢰와 불신 관계 정보가 추천 알고리즘 성능 개선에 중요한 정보가 될 수 있음을 확인하였다. 실험에 사용된 Epinions Dataset에서도 알 수 있듯이 전자상거래 환경에서는 상대방의 의견에 대한 평가를 기준으로 특정 사용자가 신뢰하거나 혹은 불신하는 대상을 쉽게 추정할 수 있다. 최근 전자상거래 사이트들의 플랫폼은 사용자들이 자유롭게 의견을 작성하거나 상대방의 의견의 유용성도 평가할 수 있도록 구축되어 있기 때문에, 본 연구의 제안 알고리즘을 적용할 수 있다는 점에서 실무적인 가치가 상당히 높을 것으로 기대된다.

둘째, 본 연구의 실증분석 결과는 신뢰와 불신 중사용자 간 불신 관계 정보가 협업필터링의 성능 개선에 더 큰 영향을 미친다는 점을 시사한다. 가장 우수한 결과를 산출한 SNACF의 경우를 보면, 신뢰 정보를 사용했을 때 보다 불신 정보를 사용했을 때 예측정확도가 더 크게 개선됨을 확인할 수 있다. 또한 제안 알고리즘인 SNACF-GA에서도 신뢰 정보, 그 중에서도 특히 직접 신뢰관계로부터 추출된 중심성은 사실상

전혀 도움이 안되는 것으로 나타났다. 요약하면, 신뢰보다는 불신을 크게 반영하였을 때, 예측정확도의 개선이 두드러지게 나타났다.

일반적으로 사용자들은 관대화 경향(leniency tendency)을 갖는데, 이는 특정 대상의 의견에 대하여 평가를 내릴 때, 실제로 본인이 인지하는 것과 다르게 평가 점수를 좀 더 후하게 주는 것을 뜻한다. 이에 비추어 볼 때, 웹에서 수집된 신뢰 관계 정보는 불신 관계 정보에 비하여 더 많은 오류를 갖고 있을 가능성이 높다. 실제로 본 연구의 성능을 검증하기 위해 사용된 Epinions 데이터셋의 경우에도 불신 관계 정보가 전체 100% 중 약 15%에 해당되며, 신뢰 관계 정보에 비해 약 6배 더 적게 입력된 것으로 조사되었다. 이것만 보더라도 사용자들이 불신보다는 신뢰를 표출하는 경우가 더 많다는 사실을 알 수 있다. 따라서 본 연구는 사용자간의 사회적인 관계를 추적하고 관리하는 측면에서 우호적인 관계보다는 사용자 간의 불신 관계, 즉 대립되는 관계에 좀 더 주목할 필요가 있음을 시사한다. 특히 선행연구들이 주로 신뢰 관계에만 집중하고, 오늘날 중요성이 점차 증대되고 있는 불신 관계를 추가적으로 고려한 사례가 지금까지 거의 없었는데, 본 연구의 이 같은 발견은 향후 연구자들이 불신 관계에 보다 주목할 필요가 있음을 시사한다.

셋째, 본 연구는 사회연결망분석을 통한 사용자 간 신뢰 및 불신 관계 분석 시, 직접적인 관계 뿐 아니라 간접적인 관계를 추가로 고려할 경우 보다 효과적이라는 점을 시사하고 있다. 특히 본 연구에서는 매개자가 1인인 간접관계만 고려하였음에도 불구하고, 상당히 유의미한 성과개선 효과를 얻을 수 있었는데, 매개자가 2인 이상인 간접관계까지 추가로 고려한다면 보다 나은 성

능의 향상을 기대해 볼 수 있을 것으로 예상된다.

반면 본 연구의 한계점은 다음과 같다. 첫째, 본 연구는 사용자 간 신뢰 및 불신 관계가 갖는 영향력을 최대한 반영하기 위해서, 해당 관계 정보가 충분히 존재하는 사용자들 위주로 표본을 임의 추출하여 실험에 활용하였다. 때문에, 이 과정에서 표본 추출의 오류가 포함되었을 가능성이 있으며, 이로 인해 연구 결과의 성능을 일반화하기에는 한계점이 존재한다. 따라서 일반화된 성능 향상을 확인하기 위해서는 관계 정보가 다소 부족해지더라도, 모든 데이터를 대상으로 실험을 하거나, 임의표본추출 방식을 적용해야 할 것으로 생각된다.

둘째, 현재 제안 알고리즘인 SNACF-GA의 경우, 사회연결망분석 시 대표성을 갖는 중심성 지표를 사용하고 있는데, 이처럼 요약된 지표를 사용하는 것 보다는 실제적인 네트워크 관계 정보를 있는 그대로 유사도 산출시 반영할 경우, 정확도가 한층 더 향상될 가능성이 있다(Choi et al., 2016). 따라서 직간접적인 신뢰 또는 불신 관계를 직접 탐색하여 유사도 산출시 고려하는 보다 고도화된 추천 알고리즘에 대한 후속 연구를 앞으로 진행할 필요가 있다.

## 참고문헌(References)

- Adomavicius, G., and Tuzhilin, A., "Toward the Next Generation of Recommender Systems : A Survey of the State-of-the-art and Possible Extensions," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, No. 6, 2005. pp. 734-749.

- Ahn, H., "Improvement of a Context-aware Recommender System through User's Emotional State Prediction," *Journal of Information Technology Applications & Management*, Vol. 21, No. 4, 2014, pp. 203-223.
- Ahn H., and Lee, H. Y., "A Combination Model of Multiple Artificial Intelligence Techniques Based on Genetic Algorithms for Investment Decision Support Aid : An Application to KOSPI," *The e-Business Studies*, Vol. 10, No. 1, 2009, pp. 267-288.
- Ahn, S.-M., Kim, I. H., Choi, B., Cho, Y., Kim, E., and Kim, M.-K., "Understanding the Performance of Collaborative Filtering Recommendation through Social Network Analysis," *Journal of Society for e-Business Studies*, Vol. 17, No. 2, 2012, pp. 129-147.
- Ahn, Y. Y., Han, S., Kwak, H., Moon, S., and Jeong, H., "Analysis of Topological Characteristics of Huge Online Social Network Service," *Proceeding of 16th International Conference on World Wide Web*, 2007, pp. 835-844.
- Barnes, J. A., "Class and committees in a Norwegian island parish," *Human Relations*, Vol. 7, 1954, pp. 39-58.
- Breese, J. S., Heckerman, D., and Kadie, C., "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," *Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI-98)*, San Francisco, California, 1998, pp. 43-52.
- Burke, R. "Hybrid recommender systems: Survey and experiments," *User Modeling and User-adapted Interaction*, Vol. 12, No. 4, 2002, pp. 331-370.
- Burke, R. "Hybrid web recommender systems," *The Adaptive Web*, pp. 377-408.
- Cho, Y. H. and Kim, J. K., "Application of Web Usage Mining and Product Taxonomy to Collaborative Recommendations in E-Commerce," *Expert Systems with Applications*, Vol. 26, No. 2, 2004, pp. 233-246.
- Cho, Y. H., and Bang, J., "Social Network Analysis for New Product Recommendation," *Journal of Intelligence and Information Systems*, Vol. 15, No. 4, 2009, pp. 183-200.
- Choi, S., Kwahk, K. -Y., and Ahn, H., "Enhancing Predictive Accuracy of Collaborative Filtering Algorithms using the Network Analysis of Trust Relationship among Users," *Journal of Intelligence and Information Systems*, Vol. 22, No. 3, 2016, pp. 113-127.
- Du, Y., Du, X., and Huang, L. "Improve the Collaborative Filtering Recommender System Performance by Trust Network Construction," *Chinese Journal of Electronics*, Vol. 25, No. 3, 2016, pp. 418-423.
- Ganu, G., Elhadad, N., and Marian, A. "Beyond the Stars: Improving Rating Predictions using Review Text Content," *WebDB*, 2009.
- 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, pp. 93-104.
- Ha, I., User Modeling-based Recommender System with Link Attributes of Trust Network, Ph.D. Dissertation, Inha University, 2014.
- Hahn, K. H., and Kim, J., "The effect of offline brand trust and perceived internet confidence

- on online shopping intention in the integrated multi-channel context," *International Journal of Retail and Distribution Management*, Vol. 37, 2009, pp. 126-141.
- Hong, S. -H., and Shin, K. -S., "Using GA based Input Selection Method for Artificial Neural Network Modeling: Application to Bankruptcy Prediction," *Journal of Intelligence and Information Systems*, Vol. 9, No. 1, 2003, pp. 227-249.
- Hurlye, R. F., *The Decision to Trust: How Leaders Create High-trust Organization*, San francisco, 2011, CA: Jossey-Bass.
- Jeon, B., and Ahn, H., "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, pp. 1-18.
- Jeong, J.-H., and Kim, J.-W., "Collaborative Filtering Techniques Using Social Network Analysis for UCC Recommendation", *Journal of Intelligence and Information Systems*, Vol. 11, No. 1, 2013, pp. 185-195.
- Kim, Y. H., *Social Network Analysis*, Parkyoungsa Publishing, 2003.
- 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, pp. 65-78.
- Ling, G., Lyu, M. R., and King, I. "Ratings meet reviews, a combined approach to recommend," *Proceedings of the 8th ACM Conference on Recommender Systems*, 2014, pp. 105-112.
- Liu, F., and Lee, H. J., "Use of social network information to enhance collaborative filtering performance," *Expert Systems with Applications*, Vol. 37, No. 7, 2010, pp. 4772-4778.
- Lu, Y., Yang, S., Chau, P. Y. K., and Can, Y., "Dynamics between the trust transfer process and intention to use mobile payment services: A cross-environment perspective," *Information & Management*, Vol. 48, No. 8, pp. 393-403, 2011.
- Massa, P., Avesani, P.: "Trust-aware Collaborative Filtering for Recommender Systems," *Proceedings of Federated Int. Conference on the Move to Meaningful Internet*, 2004, pp. 492-508.
- Mayer, R. C., Davis, J. H., and Schoorman, F. D., "An Integrative Model of Organizational Trust," *Academy of Management Review*, Vol. 30, 1995, pp. 709-734.
- McAuley, J. and Leskovec, J. "Hidden factors and hidden topics: Understanding rating dimensions with review text," *Proceedings of the 7th ACM Conference on Recommender Systems*, 2013, pp. 165-172.
- McNee, S. M., Kapoor, N., and Konstan, J. A., "Don't look stupid: avoiding pitfalls when recommending research papers," *Proceedings of the 20th Anniversary Conference on Computer Supported Cooperative Work*, 2006, pp. 171-180.
- Mislove, A., Marcon, M., Gummadi, K. P., Druschel, P., and Bhattacharjee, B., "Measurement and Analysis of Online Social Networks," *Proceedings of the 7th ACM*



- SIGCOMM Conference on Internet Measurement*, 2007, pp. 29-42.
- Noh, H., and Ahn, H. "A Study on the Recommendation Algorithm based on Trust/Distrust Relationship Network Analysis," *Journal of Information Technology Application & Management*, Vol. 24, No. 1, 2017, pp. 169-185.
- Ryu, Y. U., Kim, H. K., Cho, Y. H., and Kim, J. K., "Peer-oriented content recommendation in a social network," *Proceedings of the Sixteenth Workshop on Information Technologies and Systems*, 2006, pp. 115-120.
- Sarwar, B., Karypis, G., Konstan, J., and Riedl, J., "Application of dimensionality reduction in recommender systems: A case study," *Proceedings of the WebKDD Workshop at the ACM SIGKDD*, 2000.
- Sarwar, B., Karypis G., Konstan, J., and Riedl, J., "Item-based collaborative filtering recommendation algorithms," *Proceeding of the 10th International Conference on World Wide Web*, 2001, pp. 285-295.
- Schafer, J. B., Konstan, J. A., and Riedl, J., "E-commerce Recommendation Applications," *Data Mining and Knowledge Discovery*, Vol.5, Nos. 1-2, 2001, pp. 115-153.
- Shin, C.-H., Lee, J.-W., Yang, H.-N., and Choi, I. Y., "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, pp. 19-42.
- Son, J., Kim, S. B., Kim, H., and Cho, S. "Review and Analysis of Recommender Systems," *Journal of the Korean Institute of Industrial Engineers*, Vol. 41, No. 2, 2015, pp. 185-208.
- Tang, J., Aggarwal, C., and Liu, H., "Recommendations in signed social networks," *Proceedings of the 25th International Conference on World Wide Web*, 2016, pp. 31-40.
- Thelwall, M., "Social Networks, Gender, and Friending: An Analysis of MySpace Member Profiles," *Journal of the American Society for Information Science and Technology*, Vol. 59, No. 8, 2008, pp. 1321-1330.
- Vozalis, M., and Margaritis, K. G. "Collaborative filtering enhanced by demographic correlation," *AAAI Symposium on Professional Practice in AI, of the 18th World Computer Congress*, 2004.
- Wang, P., Gonzalez, M. C., Hidalgo, C. A., and Barabasi, A. L., "Understanding the Spreading Patterns of Mobile Phone Viruses," *Science*, Vol. 324, 2009, pp. 1071-1076.
- Wei, C., Richard K., and Simon F., "Web 2.0 Recommendation service by multi-collaborative filtering trust network algorithm," *Information Systems Frontiers*, Vol. 15, No. 4, 2013, pp. 533-551.
- Yuan, W., Improved Trust-Aware Recommender System using Small-Worldness of Trust Networks, Ph.D. Dissertation, Kyung Hee University, 2010.

Abstract

## Social Network-based Hybrid Collaborative Filtering using Genetic Algorithms

Heeryong Noh\* · Seulbi Choi\*\* · Hyunchul Ahn\*\*\*

Collaborative filtering (CF) algorithm has been popularly used for implementing recommender systems. Until now, there have been many prior studies to improve the accuracy of CF. Among them, some recent studies adopt 'hybrid recommendation approach', which enhances the performance of conventional CF by using additional information. In this research, we propose a new hybrid recommender system which fuses CF and the results from the social network analysis on trust and distrust relationship networks among users to enhance prediction accuracy. The proposed algorithm of our study is based on memory-based CF. But, when calculating the similarity between users in CF, our proposed algorithm considers not only the correlation of the users' numeric rating patterns, but also the users' in-degree centrality values derived from trust and distrust relationship networks. In specific, it is designed to amplify the similarity between a target user and his or her neighbor when the neighbor has higher in-degree centrality in the trust relationship network. Also, it attenuates the similarity between a target user and his or her neighbor when the neighbor has higher in-degree centrality in the distrust relationship network. Our proposed algorithm considers four (4) types of user relationships - direct trust, indirect trust, direct distrust, and indirect distrust - in total. And, it uses four adjusting coefficients, which adjusts the level of amplification / attenuation for in-degree centrality values derived from direct / indirect trust and distrust relationship networks. To determine optimal adjusting coefficients, genetic algorithms (GA) has been adopted. Under this background, we named our proposed algorithm as SNACF-GA (Social Network Analysis - based CF using GA).

To validate the performance of the SNACF-GA, we used a real-world data set which is called 'Extended Epinions dataset' provided by 'trustlet.org'. It is the data set contains user responses (rating scores and reviews) after purchasing specific items (e.g. car, movie, music, book) as well as trust / distrust

---

\* Graduate School of Business IT, Kookmin University

\*\* 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, 02707 Korea

Tel: +82-10-3227-7310, Fax: +82-2-910-4017, E-mail: hcahn@kookmin.ac.kr

relationship information indicating whom to trust or distrust between users. The experimental system was basically developed using Microsoft Visual Basic for Applications (VBA), but we also used UCINET 6 for calculating the in-degree centrality of trust / distrust relationship networks. In addition, we used Palisade Software's Evolver, which is a commercial software implements genetic algorithm. To examine the effectiveness of our proposed system more precisely, we adopted two comparison models. The first comparison model is conventional CF. It only uses users' explicit numeric ratings when calculating the similarities between users. That is, it does not consider trust / distrust relationship between users at all. The second comparison model is SNACF (Social Network Analysis - based CF). SNACF differs from the proposed algorithm SNACF-GA in that it considers only direct trust / distrust relationships. It also does not use GA optimization. The performances of the proposed algorithm and comparison models were evaluated by using average MAE (mean absolute error).

Experimental result showed that the optimal adjusting coefficients for direct trust, indirect trust, direct distrust, indirect distrust were 0, 1.4287, 1.5, 0.4615 each. This implies that distrust relationships between users are more important than trust ones in recommender systems. From the perspective of recommendation accuracy, SNACF-GA (Avg. MAE = 0.111943), the proposed algorithm which reflects both direct and indirect trust / distrust relationships information, was found to greatly outperform a conventional CF (Avg. MAE = 0.112638). Also, the algorithm showed better recommendation accuracy than the SNACF (Avg. MAE = 0.112209). To confirm whether these differences are statistically significant or not, we applied paired samples t-test. The results from the paired samples t-test presented that the difference between SNACF-GA and conventional CF was statistical significant at the 1% significance level, and the difference between SNACF-GA and SNACF was statistical significant at the 5%.

Our study found that the trust/distrust relationship can be important information for improving performance of recommendation algorithms. Especially, distrust relationship information was found to have a greater impact on the performance improvement of CF. This implies that we need to have more attention on distrust (negative) relationships rather than trust (positive) ones when tracking and managing social relationships between users.

**Key Words** : Recommender System, Collaborative Filtering, Trust, Distrust, Social Network Analysis, In-degree Centrality, Genetic Algorithm

Received : March 14, 2017 Revised : March 14, 2017 Accepted : March 20, 2017

Publication Type : Conference Awarded Paper Corresponding Author : Hyunchul Ahn

## 저 자 소개



### 노희룡

국민대학교 비즈니스IT전문대학원에서 비즈니스IT 전공으로 박사학위를 취득하고, 현재 찬영씨앤씨 이사로 재직 중이다. 주요 경력으로는 ASIANA IDT, POSCO ICT, POSCO 등 에서 Project Manager, Engineer로 근무하였다. 주요 관심분야는 추천시스템, 사회연결망분석, 디지털해독 등이다.



### 최슬비

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



### 안현철

현재 국민대학교 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.