

## 구조적 공백과 협업필터링을 이용한 추천시스템\*

김민건

동국대학교\_서울 일반대학원 경영정보학과  
(E-mail: kytsky@naver.com)

김경재

동국대학교\_서울 경영대학 경영학부  
(E-mail: kjkim@dongguk.edu)

본 연구에서는 사회연결망분석기법 중 하나인 구조적 공백 분석 결과를 이용하여 추천과정에 사용자의 정성적이고 감성적인 정보를 반영할 수 있는 협업필터링 기반의 추천시스템을 제안한다. 협업필터링은 추천기술 중 가장 많이 활용되고 있지만 전통적으로 확장성과 희박성 등의 문제점뿐 만 아니라 사용자-상품 매트릭스의 선호도만을 이용하여 추천을 함으로써 사용자의 정성적이고 감성적인 정보를 추천과정에 반영하지 못한다는 한계점이 있다. 본 연구에서 제안하는 추천시스템은 사회연결망분석에서 중심성 분석과 함께 연결망 내의 주요개체를 탐지할 수 있는 구조적 공백 분석을 이용하여 연결망 내의 대표 사용자들을 추출한 후 이들을 중심으로 군집을 형성한 후 각 군집색인 협업필터링을 수행하는 과정을 통해 전통적인 협업필터링에서 반영하지 못했던 정성적, 감성적 정보를 반영한다. 한편, 군집색인 협업필터링을 수행함으로써 추천의 효율성을 높일 수 있는 장점도 있다. 본 연구에서는 실제 사용자들의 상품에 대한 선호도 평가점수와 사용자들의 사회연결망 정보를 수집하여 실험을 수행하고 전통적인 협업필터링과 다양한 형태의 협업필터링과의 추천성과 비교를 통하여 제안하는 시스템의 유용성을 확인한다. 비교모형으로는 전통적인 협업필터링, 임의 군집색인 기반 협업필터링, k 평균 군집색인 기반 협업필터링을 이용한 추천시스템이며, 실험 결과, 제안한 모형이 다른 비교모형에 비해 추천성과의 정확도가 가장 우수하였다. 추천성과의 차이에 대한 통계적 유의성 검정 결과, 제안 모형은 전통적인 협업필터링 기반의 추천시스템과는 통계적으로 유의한 성과 차이가 없었으나, 다른 두 모형에 대해서는 통계적으로 유의한 성과의 차이가 있는 것으로 나타났다.

주제어 : 구조적 공백; 협업필터링; 군집색인; 사회연결망분석; 추천시스템

논문접수일 : 2014년 11월 10일    논문수정일 : 2014년 12월 9일    게재확정일 : 2014년 12월 10일  
투고유형 : 국문급행    교신저자 : 김경재

### 1. 서론

고객의 기호와 성향을 이해하고 활용하는 것은 모든 기업에게 반드시 필요한 과업이다. 특히 온라인 기업의 경우, 많은 고객들과 상품정보를 보유하고 있기에 개별 고객의 기호와 성향을 이해하고 맞춤형의 서비스를 제공하는 것이 기업 경쟁력의 핵심이 될 수 있다. 이러한 필요에 따

라 많은 온라인 기업들이 추천시스템을 활용하고 있으며, 이는 고객의 정보과다현상을 완화하여 고객의 구매 의사결정을 지원하고, 기업에 대한 친밀도를 제고하는 효과가 있다. 추천의 유용성과 필요성은 많은 기업들이 인지하고 있으나 정교한 추천이 이루어지지 않을 경우, 오히려 고객의 불만사유가 되어 영업기회를 잃을 수도 있다. 따라서 고객에게 만족스러운 추천서비스를

\* 이 논문은 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임.  
(NRF-2010-0025689, NRF-2014R1A1A2057825)

제공하기 위해서는 정교한 추천기술을 확보하는 것이 매우 중요하다. 실무와 학계에서 유용성과 정교성에서 가장 성공적인 평가를 받고 있는 추천기술은 협업필터링 (collaborative filtering; CF)이다.

가장 일반적인 형태의 협업필터링인 사용자 기반 협업필터링 (user-based CF)은 추천을 받고자 하는 고객이 선호도 평가를 한 상품에 대해 평가를 한 다른 고객들 중 선호도가 유사한 고객들이 선호하는 상품을 추천하는 기법이다. 즉, 추천의 기본 알고리즘은 다른 고객들과의 선호도 유사성을 계산한 후 유사도 높은 고객들의 선호 상품을 가중합계하여 선호도를 산출한 후 상위 선호도 상품을 추천하는 방식으로 일반적인 마케팅에서의 구전효과를 알고리즘 형태로 모사한 것이다. 협업필터링의 성공적인 활용에도 불구하고, 몇 가지 한계점들이 알려져 있는데, 대표적인 것이 확장성과 희박성 문제이다 (Kim and Kim, 2013). 확장성 문제는 선호도 평가를 하는 사용자와 선호도 평가를 받은 상품의 숫자가 늘어 남에 따라 고객 간 유사도 계산이나 선호 상품 추출에 소요되는 시간이 기하급수적으로 증가하는 현상을 말한다. 인터넷 쇼핑물과 같이 신속한 추천결과를 제시하여야 하는 환경에서는 치명적인 한계점이며, 이 한계점을 극복하기 위해 상품기반 협업필터링이나 하이브리드 추천기법 등이 제시되어 왔다. 희박성 문제는 신규 고객이나 신규 상품이 유입됨에 따라 선호도 평가를 하지 않은 고객이나 평가를 받지 않은 상품에 대한 유사도 계산과 추천이 불가능한 한계점을 의미한다. 이 경우에 대해 선행연구에서는 내용기반추천기법이나, 상품연결망 정보를 활용한 보완방법을 제안하여 왔다.

본 연구에서는 상기의 두 가지 한계점 외에 또

다른 한계점을 제시하고자 하는데, 이는 기존 협업필터링이 고객의 정성적이고 감성적인 정보를 반영하지 못한다는 점이다. 일반적으로 협업필터링은 사용자-상품 매트릭스 상의 선호도 평가 점수만을 이용하여 추천함으로써 사용자 간의 친분관계나 인구통계학적 정보의 유사성을 전혀 고려하지 못한다는 단점이 있다. 구체적으로 설명하면, 선호도의 유사성을 계산한 후 추출된 유사한 고객 중 본인과 친분이 있거나 인구통계학적으로 유사성이 높은 경우에도 이에 대한 고려를 따로 하지 않는다. 그러나 Shinha and Swearingen (2001), Liu and Lee (2010), Kim and Ahn (2010) 등에 따르면 사용자들은 일반적으로 본인과 친분이 있는 사용자들의 추천을 더 신뢰하고 선호하는 것으로 알려져 있다. 이는 사용자 간의 친분 관계가 사용자의 정성적이고 감성적인 유사성과 신뢰성을 제고하는 효과가 있기 때문인 것으로 판단된다. 본 연구에서는 협업필터링의 추천과정에서 이러한 사용자의 정성적이고 감성적인 정보가 반영될 수 있도록 사회연결망의 정보를 반영하는 추천시스템을 제안한다.

제안하는 추천시스템에서 사용자의 사회연결망 정보를 반영하기 위하여 먼저 사용자들의 사회연결망 정보를 수집한 후 사회연결망에서의 사결정의 주도자 또는 관계 형성의 주도자인 핵심인물을 추출하기 위한 구조적 공백 분석을 실시하고, 이에 따라 추출된 구조적 공백에 해당하는 사용자를 초기 중심점으로 하는 k평균 군집 분석을 수행한다. 이후 군집분석을 통해 얻어진 군집색인을 이용하여 군집색인 협업필터링을 수행한다. 이를 통해 사용자와 친분관계에 있는 사용자와 그와 유사한 사용자의 정보를 추천과정에 반영할 수 있다.

본 연구에서는 실제 추천성과의 개선 정도를

확인하기 위하여 실제 사용자의 선호도 평가정보와 사회연결망 정보를 수집하였으며, 이를 통해 전통적인 협업필터링, 임의 군집색인 협업필터링, k평균 군집분석 기반 군집색인 협업필터링 등과 추천성적을 비교할 것이다.

본 연구는 다음과 같이 구성되었다. 2장에서는 사회연결망분석과 추천시스템에 관한 선행연구들을 살펴보고, 3장에서는 제안하는 협업필터링과 구조적 공백을 이용한 추천시스템을 설명한다. 4장에서는 실제 사용자의 선호도 평가점수자료와 사회연결망 자료를 활용하여 제안시스템을 실증분석한 후 제안시스템의 실험결과를 제시하고 이를 다른 비교모형들의 추천성과 비교한다. 마지막으로 5장에서는 본 연구의 결론 및 향후 연구방향에 대해 설명한다.

## 2. 선행연구

본 장에서는 사회연결망분석과 추천시스템에 관한 선행연구를 살펴 보고, 구조적 공백에 관련된 이론적 배경을 설명한다.

### 2.1. 사회연결망분석과 추천시스템

사회연결망은 개인적인 인간관계가 확산되어 형성되는 사람들 사이의 연결된 네트워크를 의미한다 (Sohn, 2002; Park et al., 2009). 사회연결망분석은 이러한 사회연결망을 분석하는 방법으로 구조나 연결망 형태의 특징을 추출하고, 관계성으로 체계의 특성을 설명하거나 체계를 구성하는 단위의 행위를 설명하는 것이다 (Kim, 2007). 사회연결망분석은 사회학과 사회심리학 연구진에 의해 방법론이 처음 개발되었으며 최

근에는 통계학, 컴퓨터과학, 경제학, 마케팅, 산업공학 등 여러 분야에서 활용되고 있다. 사회연결망분석은 조직, 팀과 같은 여러 사회적 구조의 행동을 설명하고, 개인적인 인간관계가 확산되어 사람과 사람들 사이에 연결된 연결망을 보여 주는데 사용되어 왔다 (Kukkonen et al., 2010; Sohn, 2002; Park et al., 2009).

최근의 추천시스템 연구에서도 사회연결망의 정보나 사회연결망분석의 결과를 추천과정에 활용하기 위한 연구가 진행되어 왔다. 초기의 연구들은 주로 단순히 사회연결망의 정보를 추천과정에 반영하는 것이 많았으며, 사회연결망분석을 본격적으로 사용하지 않았다. Golbeck (2006)은 웹 기반의 사회네트워크 정보를 영화 추천시스템에 접목한 Film Trust라는 추천시스템을 제안하였다. 그녀의 연구에서는 평균과 비교하여 다양한 평가 의견이 제시되었을 때 더 정확한 추천이 가능함을 보여 주었다. Jyun-Cheng and Chui-Chen (2008)은 온라인 경매에서 판매자의 신뢰관계에 기반한 거래관계 추천시스템을 제안하였다. 그들의 연구에서는 두 개의 사회연결망 지표를 이용하여 협업기반의 추천시스템을 제안하였으며, 이 시스템을 통해 76%의 불량사용자 계정을 찾아 낼 수 있었다고 주장하였다. Yang and Dia (2008)는 타겟광고를 생성하는데 있어서 사회연결망의 정보를 이용하였다. 그들의 연구에서는 고객 간의 상호작용 자료를 이용하여 결속력이 강한 세부집단을 발견하고 이를 이용하여 고객의 상품 분류 별 선호확률을 추정하였으며, 이를 통해 타겟광고를 생성하는 시스템을 구축하였다. Debnath et al. (2008)은 내용 기반 추천시스템과 협업필터링을 결합한 추천시스템을 제안하였는데, 이때 내용기반 추천에 사용된 속성들의 가중치를 사회연결망 그래프로부터 얻어

진 선형 회귀분석 결과를 통해 추정하였다. 또한, Liu and Lee (2010)는 사회연결망 서비스에서의 친구관계 정보를 이용하여 협업필터링에서 친구관계에 있는 사용자의 선호도 정보를 포함하거나 가중하여 반영하는 방법을 제안하였다. 이를 통해 친구관계에 있는 사용자의 선호도가 추천 과정에 보다 더 많이 반영되게 하는 것으로, 제안하는 방법에 의한 추천성고가 기존의 방법에 비해 개선되는 것으로 나타났다. 전술한 연구들은 사회연결망의 정보를 추천과정에 반영하고자 하였으나, 사회연결망분석을 실제 활용한 것은 아니고 실증분석 없는 탐색적인 연구이거나 단순히 사회연결망의 정보만을 추천과정에 반영한 연구들이었다.

비교적 최근의 연구들은 사회연결망 정보를 단순히 활용하는 것이 아니라 사회연결망분석을 수행하여 그 결과물을 추천과정에 이용하는 연구들이 진행되었다. Park et al. (2009)은 사회연결망분석의 중심성 개념을 상품구매 네트워크에 적용하여 추천상품을 추출하는 시스템을 제안하였다. 그들의 연구에서는 공통적인 상품을 구매하는 사용자들을 연결하는 고객 연결망을 구축한 후, 중심성 값이 높은 상위 사용자의 구매목록을 이용하여 다른 사용자에게 상품을 추천하는 방식을 제안하였다. 또한, Cho and Bang (2009)는 동일한 고객들에 의해 구매되는 상품들의 연결망을 구축한 후, 사회연결망분석의 중심성 개념을 이용하여 상품 간의 구매관계를 파악한 후, 이를 기반으로 신상품을 구매할 가능성이 높은 사용자를 추출하여 신상품을 추천하는 방식을 제안하였다. 한편 Kang (2010)은 고객 연결망을 구축한 후 사회연결망분석의 구조적 공백 위치에 있는 고객들을 추출한 후 이들이 구매한 상품을 신규고객에게 추천하는 방법을 제안하였

으며, Cho and Bang (2011)은 Cho and Bang (2009)가 유사한 방법이지만 연결, 근접, 매개중심성을 가중치로 결합한 결합 중심성 지수를 이용한 방법을 제안하였다. 한편, Kang (2011)은 상품 연결망에 구조적 공백 분석을 적용하여 신상품 추천에 활용하였다.

상기의 네 연구는 사회연결망분석의 기법을 상품 네트워크에 적용한 것이다. 이들 연구에서는 희박성의 문제가 발생하는 신상품과 신규 고객에 대한 추천을 할 수 있는 방법으로 제시되었으나, 많은 사용자들이 구매하거나 많은 상품을 구매하는 중심성 높은 상품이나 사용자를 이용하여 추천하는 것이므로, 개별 사용자의 선호도를 정교하게 고려하여 추천하는 방식은 아니었다.

한편, 최근의 연구들은 일반적인 사용자의 선호도 점수를 이용하는 일반적인 협업필터링에 사회연결망분석의 결과를 통합하여 활용하고자 하고 있다. Kim and Ahn (2010)은 사회연결망분석의 연결중심성지표를 계산한 후 상위 사용자 5명을 추출하여 이들을 초기 군집중심점으로 하는 군집색인 협업필터링을 이용한 추천시스템을 제안하였다. 그들의 제안 시스템은 사회연결망 분석에서 매우 중요한 지표로 알려져 있는 연결 중심성분석의 결과물을 이용하여 군집색인 협업필터링을 수행함으로써 사회연결망분석과 협업필터링을 유기적으로 통합한 추천시스템을 제안하였다는 의의가 있다.

## 2.2. 구조적 공백

사회연결망분석은 다양한 기법들로 구성되어 있다. 사회연결망분석 기법에서 연결망의 관계 구조를 파악하기 위한 방법으로 가장 많이 사용

되는 지표는 중심성(연결정도중심성, 매개중심성, 근접중심성), 집중도(centralization), 구조적 공백(structural hole), 밀도(density), 포괄성(inclusiveness) 등이 있다 (Bonacich, 1987; Sohn, 2002; Kim, 2007). 이 중 선행연구에서 추천시스템에 활용된 지표는 중심성과 구조적 공백이다. 중심성은 비교적 많은 연구에서 활용되어서 널리 알려져 있으나, 구조적 공백은 상대적으로 적게 활용되어 왔다.

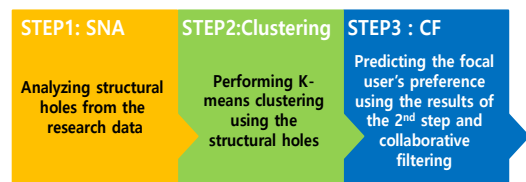
구조적 공백 개념은 Burt (1992)에 의해서 처음 개념이 정의되었는데, ‘사회적 구조 내에 존재하는 빈 공간’으로 정의된다. 즉, Burt (1992)에 따르면, 한 고객이 여러 수많은 고객들과 비중복적인 관계를 맺을 수 있는지에 관련된 개념으로 구조적 공백이 없는 네트워크와 구조적 공백이 존재하는 네트워크로 비교 할 수 있다. 일반적으로 구조적 공백에 위치한 개체는 사회연결망 내의 정보흐름에 중요한 영향력을 미치는 것으로 알려져 있다 (Rodan and Galunic, 2004; Rodan, 2010). 또한, 구조적 공백에 해당하는 개체는 주변 네트워크 개체들에게 있어서 상대적인 영향력을 가지는 개체로 인식된다 (Burt, 1992).

일반적으로 구조적 공백을 측정하는 직접적인 단일 지표는 존재하지 않지만, 네트워크 제약의 측정은 효율성(efficiency), 적합한 규모(effective size), 제약성(constraint), 계층(hierarchy)을 통해 파악할 수 있다. 적합한 규모는 각 개체가 가지고 있는 연결관계의 수에서 상대방이 가지고 있는 연결관계의 수의 평균을 제외한 값으로 내가 관계를 맺고 있는 상대방이 다른 사람들과 적은 수의 관계를 유지할 경우 나의 네트워크의 효과가 커진다는 사실을 기반으로 한다. 효율성은 적합한 규모를 실제 네트워크 규모로 나누어 구하며, 이를 통해 표준화된 네트워크의 적합한 규모

를 구할 수 있다. 제약성은 각 개체의 제약성을 의미하는데 이 값이 낮을수록 네트워크에서 유리한 위치에 있음을 의미한다. 따라서, 제약성이 높다는 의미는 다른 구성원과 관계를 지속하는 과정에서 중개인의 영향이 강하기 때문에 관계에 제약성이 발생하고, 구조적 공백까지 발생될 가능성이 생긴다는 의미이다. 마지막으로 계층은 각 개체들의 제약성이 얼마나 특정 개체에 집중되어 있는지를 측정하는 지표이다 (Burt, 1992; Sohn, 2002; Kim, 2007; Hyun et al., 2011).

### 3. 구조적 공백과 협업필터링을 이용한 추천시스템

이 장에서는 본 연구에서 제안하는 협업필터링과 구조적 공백을 이용한 추천시스템의 수행 절차를 설명한다. 특히 본 연구에서는 협업필터링을 수행하는 과정에서 사용자의 사회연결망 정보가 반영될 수 있도록 사회연결망분석 기반의 군집색인 협업필터링을 수행하게 된다. 본 연구에서 일반적인 협업필터링 대신 군집색인 협업필터링을 활용하는 이유는 사회연결망의 대표사용자들의 특성을 협업필터링에 보다 더 많이 반영할 수 있는 장점과 일반적인 협업필터링에 비해 연산량이 적어서 효율성이 높기 때문이다. 본 연구에서 제안하는 추천시스템의 수행과정은



〈Figure 1〉 The process of the proposed recommender systems

<Figure 1>과 같으며 크게 아래의 세 단계로 구성된다.

#### - 구조적 공백 분석 단계

첫 번째 단계에서는 사용자의 사회연결망 정보를 활용하여 구조적 공백 분석을 수행하여 사회연결망을 대표할 수 있는 k명의 대표 사용자들을 추출한다. 대표 사용자들은 사회연결망 내의 여러 사용자들의 특성을 대표하는 사용자들로서 이들의 특성을 이용하여 사용자들을 몇 개의 군집으로 집단화할 수 있다.

#### - 군집화 단계

다음으로는 구조적 공백 분석단계에서 도출된 k명의 대표 사용자들의 특성을 이용하여 사회연결망 내의 사용자들을 비슷한 속성을 가진 여러 개의 집단으로 군집화한다. 군집화 과정은 가장 일반적인 k평균 군집분석을 이용하며 구체적으로는 k평균 군집분석의 초기 군집 중심점으로 대표 사용자들의 선호도 벡터를 이용하여 군집화를 수행한다. 여기서 사용자 간의 유사도는 상품에 대한 선호도 벡터의 유사성을 이용하여 계산하는데, 일반적으로 상품에 대한 선호도 벡터는 사용자가 모든 상품에 대한 평가를 할 수 없으므로 일반적인 유클리드 거리 방식으로 계산하기 곤란하다. 따라서 본 연구에서는 유클리드 거리 대신 피어슨 상관계수를 이용하여 상품 간 선호도 벡터의 유사도를 계산하는 방식을 사용한다.

#### - 협업필터링 단계

마지막 단계에서는 협업필터링을 수행하여 추천대상 사용자와 선호도가 유사한 이웃 사용자

들의 정보를 이용하여 추천을 하게 된다. 일반적인 협업필터링과는 달리 구조적 공백 분석 기반 군집색인 방법에서는 두 번째 단계를 통해 추천대상 사용자와 같은 군집에 소속된 사용자들의 선호도 평가점수에 가중치를 부여하여 그들의 선호도가 추천과정에 더 많이 반영되도록 한다. 이는 추천대상 사용자와 전혀 관계가 없는 사용자들보다는 사회적으로 연결관계에 있는 사용자들이 추천대상 사용자와 보다 많은 정서적 공감대를 가질 것이라는 가정에 기반한 것이다. 사회연결망의 정보를 추천과정에 반영하기 위해서는 단순히 일반적인 협업필터링의 이웃 사용자를 사회연결망 상의 친구 중에서 선호도가 유사한 사용자로 대체하면 된다. 하지만, 일반적으로 다른 사용자에 비해 사용자의 사회연결망 상의 친구 중 선호도까지 유사한 친구는 극히 소수인 경우가 많으므로 사회연결망의 정보를 추천과정에 반영하기에는 비효과적이다 (Liu and Lee, 2010). 따라서 본 연구에서는 Liu and Lee (2010)과 Kim and Ahn (2010)에서 제안한 방법을 활용한다. 즉, 협업필터링의 이웃 사용자로 선정된 사용자 중에서 동일한 군집색인을 가지는 경우에는 가중치(예를 들면 두 배의 가중치)를 부여하고 나머지 사용자에 대해서는 가중치를 부여하지 않아서 사회연결망 상의 친구의 선호도가 보다 더 추천과정에 반영될 수 있도록 하는 것이다.

## 4. 연구데이터와 실험결과

### 4.1. 연구데이터 수집

본 연구에서 제안하는 추천시스템의 추천성공률을 확인하기 위하여 실제 사용자들의 데이터를

수집하여 실험을 수행한다. 수집된 연구데이터는 사용자들의 상품 선호도 평가점수자료와 사회연결망 관계에 대한 자료이다. 일반적으로 추천시스템 연구에서는 MovieLens 데이터나 EachMovie 데이터를 활용하여 추천성적을 평가하지만, 이들 데이터에는 사용자의 사회연결망 자료가 포함되어 있지 않으므로 직접 연구데이터를 수집하였다. 연구데이터 수집을 위해 웹 기반의 연구데이터 수집 시스템을 구축하였으며, 사용자의 사회연결망 정보와 관련하여서는 페이스북, 마이스페이스, 싸이월드와 트위터 등의 마이크로 블로그 서비스에서의 친구관계 자료를 수집하였다. 상품 선호도 평가와 관련하여서는 일반적으로 추천시스템 연구에서 많이 사용해진 영화에 대한 평가점수 자료를 수집하였다. 그 밖에 사용자의 기초 인구통계자료를 수집하였다. 수집 대상은 국내 대학의 재학생으로 한정하였으며, 총 91명의 자료를 수집하였으나 불성실한 답변이 포함된 하나의 자료를 제외한 90명의 자료를 연구데이터로 활용하였고 선호도 평가 대상 영화는 100편으로 한정하여 평가를 실시하였다.

#### 4.2. 실험설계와 결과

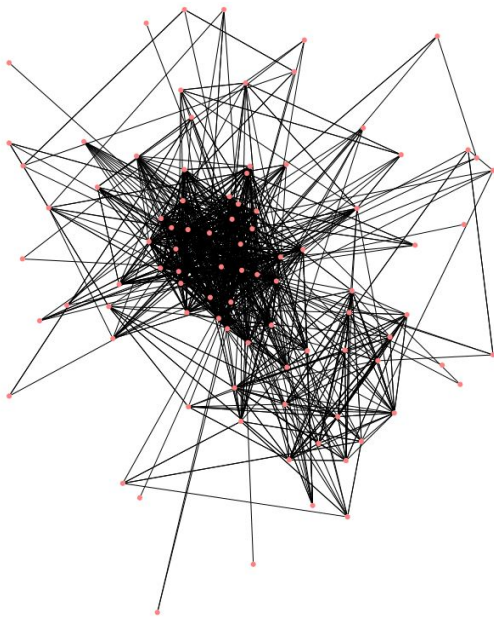
본 연구에서 제안하는 추천시스템의 유용성 확인을 위하여 두 가지 독립된 실험과정을 진행하였다. 첫 번째 실험은 구조적 공백 분석으로 이를 위해 UCINET 6를 이용하였다. UCINET 6는 사회연결망분석을 위한 대표적인 분석 프로그램으로 Lin Freeman에 의해 개발되었으며, 학술연구에 널리 이용되고 있다. 두 번째 실험은 k평균 군집색인 협업필터링을 수행하였다. 이 실험 환경은 Microsoft Excel의 VBA (Visual Basic

Applications)를 이용하여 개발하였으며, k평균 군집분석과 협업필터링이 통합되어 실행될 수 있도록 하였다. 전 장에서 설명한 바와 같이 본 연구에서 사용한 k평균 군집분석은 일반적인 k평균 군집분석과 달리 사용자의 선호도 평가점수 벡터를 기반으로 유사성을 평가하여야 하므로 유사성 평가방법은 선호도 평가점수 벡터의 피어슨 상관계수를 이용하였다.

한편, 본 연구에서 제안하는 추천시스템의 성과를 보다 객관적으로 평가하기 위해 본 연구에서는 여러 추천시스템과의 비교평가를 수행하였다. 비교평가에 활용된 추천시스템 중 첫 번째는 일반적인 협업필터링을 이용한 추천시스템으로 본 연구에서는 CCF (Conventional CF)라 표기하였다. 또 다른 비교모형은 일반적인 k평균 군집분석 기반의 협업필터링을 활용하는 추천시스템으로 KCF (k-means cluster-indexing CF)로 표기하였다. CCF는 사용자의 상품 선호도 평가점수 이외에 다른 정보를 전혀 활용하지 않는 일반적인 협업필터링 모형이고, KCF는 일반적인 군집색인 협업필터링 모형으로 SCF와의 비교를 통해 사회연결망 정보를 추천과정에 고려하였을 경우와의 차이를 확인하기 위한 모형이다. 본 연구에서 k평균 군집분석의 군집의 개수 (k)는 5로 결정하였는데 k평균 군집분석의 경우, k를 결정할 수 있는 일반적인 방법이 없고 실험자의 판단에 따라 결정하여야 하므로 데이터 총수를 고려하여 다섯 개로 설정하였다. 세 번째 추천시스템은 임의배정 군집분석 기반의 협업필터링을 사용하는데 본 연구에서는 RCF (Randomly selected cluster-indexing CF)로 표시하였다. 이는 k평균 군집분석의 효과를 배제하기 위하여 사용자들을 다섯 개의 군집으로 임의 배정하여 군집색인을 배정한 후 협업필터링을 수행하는 것이다. 마치

막으로 본 연구에서 제안하는 구조적 공백 기반의 군집색인 협업필터링을 이용한 추천시스템으로 SCF(Structural hole-based cluster-indexing CF)로 표시하였다. 본 연구에서는 동일한 군집색인에 속하는 사용자의 선호도 점수에 대해서 두 배의 가중치를 부여하는 방식을 이용하였는데 이는 전술한 바와 같이 같은 군집색인을 가진 사용자가 소수인 경우가 많으므로 추천과정에 사회연결망 정보를 보다 많이 반영하기 위해 Liu and Lee (2010), Kim and Ahn (2010) 등의 연구에서 사용한 방식을 이용하였다.

본 연구에서 사용된 사용자들의 사회연결망은 아래 <Figure 2>와 같이 나타났다.



<Figure 2> Social network in this study

<Figure 2>의 사회연결망 자료에 대하여 UCINET 6를 이용하여 구조적 공백 분석을 수행하였으며 <Table 1>은 분석 결과 구조적 공백에

위치하고 있는 상위 사용자 5명의 결과이며, 이 사용자들이 k평균 군집분석의 초기 군집 중심점으로 사용되었다.

<Table 1> The results of the structural hole analysis

User ID	Efficiency	Effective size	Constraint	Hierarchy
u76	0.522	18.257	0.090	0.024
u19	0.508	16.758	0.094	0.021
u66	0.464	16.257	0.095	0.022
u3	0.442	14.576	0.097	0.017
u17	0.422	14.353	0.100	0.016

한편, 본 연구에서 추천시스템의 추천성과를 평가하기 위하여 평균MAE를 사용하였다. MAE (mean absolute error)는 사용자의 실제 평가와 예측 평가점수의 차이를 의미하는 것으로 협업필터링 연구에서 자주 이용되고 있다 (Breese et al., 1998; Sarwar et al., 1998; Goldberg et al., 2001). 평균MAE는 식(1)과 같이 표현된다.

$$Avg.MAE = \left\{ \sum_{k=1}^N \left( \sum_{i=1}^n |p_{ki} - a_{ki}| / n \right) \right\} / N \quad \text{식(1)}$$

식(1)에서  $N$ 은 데이터셋에서의 사용자의 수를 의미하고,  $n$ 은 데이터셋에서의 상품의 수를,  $p_{k,i}$

<Table 2> Recommendation performances of the comparative models

Model	Average MAE	Standard deviation of MAE
CCF	0.826	0.284
RCF	0.853	0.276
KCF	0.852	0.279
SCF	0.825	0.286



는 상품  $i$  에 대한 사용자  $k$  의 예측된 평가점수를,  $a_{k,i}$  는 상품  $i$  에 대한 사용자  $k$  의 실제 평가점수를 의미한다. <Table 2>는 본 연구에서 제안하는 추천시스템과 비교 추천시스템들의 추천성적을 정리한 것이다.

<Table 2>에서 제시된 바와 같이 본 연구에서 제시하는 모형인 SCF가 다른 비교모형에 비해 평균MAE가 가장 작아서 추천성적이 가장 우수한 것으로 나타났다. 같은 군집색인 협업필터링 모형인 RCF와 KCF는 SCF와 CF에 비해 추천성적이 떨어졌으며, 둘 간의 비교에서는 KCF가 우월한 것으로 나타났다. 한편, 군집색인 기반의 협업필터링인 RCF와 KCF의 추천성적이 일반적인 CF인 CCF에 비해 저조한 것으로 나타났다. 이는 군집색인 협업필터링이 연산을 감소시켜서 효율성을 제고하는 반면, 정보손실이 크기 때문에 모든 사용자를 이용하여 추천을 하는 CCF에 비해 추천성적이 저조할 수 있다는 일반적인 사실을 보여준 것이다. 한편, SCF의 경우에는 모든 모형에 비해 추천성적이 우수하였으며, 특히, CCF에 비해서도 성적이 우수하였는데, 이는 일반적인 군집색인 협업필터링의 효율성을 가지면서도 추천의 성적이 떨어지지 않았음을 보여주는 것이다.

본 연구에서는 추천성적의 차이가 얼마나 통계적으로 유의한 지 확인하기 위하여 대응표본 t-test를 수행하였으며, 그 결과는 <Table 3>와 같다.

<Table 3> The results of statistical significance test for recommendation performance (p-value)

	RCF	KCF	SCF
CCF	0.000	0.000	0.206
RCF		0.841	0.000
KCF			0.000

<Table 3>에 제시된 결과를 살펴보면, SCF의 추천성과 RCF, KCF의 추천성적의 차이가 각각 1% 유의수준에서 통계적 유의성을 가짐을 알 수 있다. 또한, CCF도 RCF, KCF의 성과와 비교하여 각각 1% 유의수준에서 유의한 성과 차이가 있다는 것을 보여준다. 추천의 성과에서 SCF가 CCF보다 우수하였으나 그 성과의 차이는 통계적으로 유의하지 않았다.

## 5. 결론과 향후 연구방향

본 연구에서는 일반적인 협업필터링과 사회연결망 분석을 통해 얻어진 사회연결망 정보를 추천과정에 반영할 수 있는 방법을 제안하였다. 실제 사용자의 데이터를 이용한 분석 결과, 제안한 추천시스템이 추천의 성과를 개선할 수 있는 것으로 나타났다. 비교모형 분석을 통해 유사한 시스템과의 성과 차이에서 확인한 것과 같이 동일한 시스템 구조에서도 사회연결망 정보가 반영될 때 추천의 성과가 보다 더 개선되는 것으로 나타났는데 이는 사회연결망 정보에 포함된 사용자의 정성적, 감성적인 정보가 반영된 것으로 생각된다.

본 연구의 제안 추천시스템의 유용성이 확인되었지만 일반적인 협업필터링에 비해 성과 개선의 정도가 유의할 정도로 크지 않다는 점과 실험실 환경에서 제한된 연구데이터를 사용하였다는 점은 본 연구의 중요한 한계점이다. 본 연구에서는 사회연결망 정보를 가중치를 주어 반영하는 방식을 제안하였는데 사회연결망 정보를 보다 직접적으로 많이 반영할 수 있는 방안이 제안된다면 보다 큰 성과 개선이 있을 것으로 기대된다. 또, 여러 가지 환경요인이 제한된 실험실

환경의 데이터 수집과정을 통해 데이터를 수집 하였기에 실제 데이터에서도 동일한 성과의 개선을 가져 올 것인지에 대한 확신을 가질 수 없다는 점도 단점이다. 그러나 이 부분은 최근에 강화된 개인정보보호법 등으로 개인정보의 수집이 더욱 어려워졌기에 연구환경에서는 어쩔 수 없는 한계점이라 할 수도 있다. 따라서 연구데이터 수집과정에서 사용자의 활용동의와 함께 보다 체계적인 데이터 수집계획 수립을 통해 실제 데이터와 유사한 데이터를 수집할 수 있어야 할 것이며, 이는 향후 연구과제가 될 것이다.

## 참고문헌(References)

- Bonacich, P., "Power and centrality : a family of measures," *American Journal of Sociology*, Vol.92, No.5(1987), 1170~1182.
- 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.
- Burt, R. S., "Structural Holes : The Social Structure of Competition," *Harvard University Press*, Cambridge MA, 1992.
- 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.
- Cho, Y. and J. Bang, "Applying centrality analysis to solve the cold-start and sparsity problems in collaborative filtering," *Journal of Intelligence and Information Systems*, Vol.17, No.3(2011), 99~114.
- Debnath, S., N. Granguly, and P. Mitra, "Feature weighting in content based recommendation system using social network analysis," *Proceedings of the 17th international conference on World Wide Web*, (2008), 1041~1042.
- 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.
- Goldberg, K., T. Roeder, D. Gupta, and C. Perkins, "Eigentaste: a constant time collaborative filtering algorithm," *Information Retrieval*, Vol.4, No.2(2001), 133~151.
- Wang, J.-C. and C.-C. Chiu, "Recommending trusted online auction sellers using social network analysis," *Expert Systems with Applications*, Vol.34, No.3(2008), 1666~1679.
- Hyun, Y. S., E. K. Shin, and H. R. Lee, "The roles of trust and broker at structural holes in multi-cultural learning community," *The Korean Journal of Lifelong Education*, Vol. 17, No. 2(2011), 1~31.
- Kang, B. - S., "A novel web recommendation method for new customers using structural holes in social networks," *Journal of Industrial Economics and Business*, Vol.23, No.5(2010), 2371~2385.
- Kang, B. - S., "Structural holes method in social networks for new product recommendations," *Journal of the Korean Data Analysis Society*, Vol.13, No.3(B)(2011), 1365~1377.
- Oinas-Kukkonen, H., K. Lyytinen, and Y. Yoo, "Social networks and information systems: ongoing and future research streams," *Journal of the Association for Information Systems*,

- Vol.11, No.2(2010), 61~68.
- 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 Y. Kim, "Recommender system using implicit trust-enhanced collaborative filtering," *Journal of Intelligence and Information Systems*, Vol.19, No.4(2013), 1~10.
- Kim, Y. H., *Social Network Analysis*, Parkyoungsa Publishing, 2007.
- 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.
- Rodan, S. and C. Galunic, "More than network structure : how knowledge heterogeneity influences managerial performance and innovativeness," *Strategic Management Journal*, Vol.25, No.6 (2004), 541~ 562.
- Rodan, S., "Structural holes and managerial performance : identifying the underlying mechanisms," *Social Networks*, Vol.32, No.3(2010), 168~179.
- Sarwar, B. M., J. A. Konstan, A. Borchers, J. Herlocker, B. Miller, and J. Riedl, "Using filtering agents to improve prediction quality in the GroupLens research collaborative filtering system," *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work(CSCW)*, (1998), 345~354.
- 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), Dublin, Ireland.
- Sohn, D. W., *Social Network Analysis*, Kyungmunsa Publishing, 2002.
- Yang, W.-S. and J. - B. Dia, "Discovering cohesive subgroups from social networks for targeted advertising," *Expert Systems with Applications*, Vol.34, No.3(2008), 2029~2038.

Abstract

## Recommender Systems using Structural Hole and Collaborative Filtering

Mingun Kim\* · Kyoung-jae Kim\*\*

This study proposes a novel recommender system using the structural hole analysis to reflect qualitative and emotional information in recommendation process. Although collaborative filtering (CF) is known as the most popular recommendation algorithm, it has some limitations including scalability and sparsity problems. The scalability problem arises when the volume of users and items become quite large. It means that CF cannot scale up due to large computation time for finding neighbors from the user-item matrix as the number of users and items increases in real-world e-commerce sites. Sparsity is a common problem of most recommender systems due to the fact that users generally evaluate only a small portion of the whole items. In addition, the cold-start problem is the special case of the sparsity problem when users or items newly added to the system with no ratings at all. When the user's preference evaluation data is sparse, two users or items are unlikely to have common ratings, and finally, CF will predict ratings using a very limited number of similar users. Moreover, it may produce biased recommendations because similarity weights may be estimated using only a small portion of rating data.

In this study, we suggest a novel limitation of the conventional CF. The limitation is that CF does not consider qualitative and emotional information about users in the recommendation process because it only utilizes user's preference scores of the user-item matrix. To address this novel limitation, this study proposes cluster-indexing CF model with the structural hole analysis for recommendations. In general, the structural hole means a location which connects two separate actors without any redundant connections in the network. The actor who occupies the structural hole can easily access to non-redundant, various and fresh information. Therefore, the actor who occupies the structural hole may be an important person in the focal network and he or she may be the

---

\* Department of Management Information Systems, Graduate School, Dongguk University\_Seoul  
\*\* Corresponding Author: Kyoung-jae Kim  
Business School, Dongguk University\_Seoul  
30 Pildong-ro 1gil, Chung-gu, Seoul 100-715, Korea  
Tel: 02-2260-3324, Fax: 02-2260-3684, E-mail: kjkim@dongguk.edu

representative person in the focal subgroup in the network. Thus, his or her characteristics may represent the general characteristics of the users in the focal subgroup. In this sense, we can distinguish friends and strangers of the focal user utilizing the structural hole analysis.

This study uses the structural hole analysis to select structural holes in subgroups as an initial seeds for a cluster analysis. First, we gather data about users' preference ratings for items and their social network information. For gathering research data, we develop a data collection system. Then, we perform structural hole analysis and find structural holes of social network. Next, we use these structural holes as cluster centroids for the clustering algorithm. Finally, this study makes recommendations using CF within user's cluster, and compare the recommendation performances of comparative models.

For implementing experiments of the proposed model, we composite the experimental results from two experiments. The first experiment is the structural hole analysis. For the first one, this study employs a software package for the analysis of social network data – UCINET version 6. The second one is for performing modified clustering, and CF using the result of the cluster analysis. We develop an experimental system using VBA (Visual Basic for Application) of Microsoft Excel 2007 for the second one. This study designs to analyzing clustering based on a novel similarity measure – Pearson correlation between user preference rating vectors for the modified clustering experiment. In addition, this study uses 'all-but-one' approach for the CF experiment.

In order to validate the effectiveness of our proposed model, we apply three comparative types of CF models to the same dataset. The experimental results show that the proposed model outperforms the other comparative models. In especial, the proposed model significantly performs better than two comparative modes with the cluster analysis from the statistical significance test. However, the difference between the proposed model and the naive model does not have statistical significance.

**Key Words** : Structural hole; Collaborative filtering; Cluster-indexing; Social network analysis;  
Recommender system

Received : November 10, 2014    Revised : December 9, 2014    Accepted : December 10, 2014

Type of Submission : Fast Track    Corresponding Author : Kyoung-jae Kim

## 저자 소개



**김민건**

현재 동국대학교 일반대학원 경영정보학과 석사과정에 재학 중이다. 동국대학교 컴퓨터 멀티미디어학부에서 공학사를 취득하였으며, *지능정보연구*에 논문을 게재하였다. 연구 관심분야는 고객관계관리, 사회연결망분석, 데이터마이닝, 소셜네트워크 서비스(SNS) 등이다.



**김경재**

현재 동국대학교 경영대학 경영학부 교수로 재직 중이다. KAIST에서 경영정보시스템을 전공으로 박사학위를 취득하였으며, *경영학연구*, *지능정보연구*, *Annals of Operations Research*, *Applied Intelligence*, *Applied Soft Computing*, *Asia Pacific Journal of Information Systems*, *Computers & Operations Research*, *Computers in Human Behavior*, *Expert Systems*, *Expert Systems with Applications*, *Information*, *Intelligent Data Analysis*, *International Journal of Electronic Commerce*, *Intelligent Systems in Accounting, Finance and Management*, *Neural Computing & Applications*, *Neurocomputing* 등의 학술지에 논문을 게재하였다. 연구 관심분야는 고객관계관리, 데이터마이닝, 비즈니스 인텔리전스, 지식경영 등이다.