소셜네트워크 분석을 통한 협업필터링 추천 성과의 이해

Understanding the Performance of Collaborative Filtering Recommendation through Social Network Analysis

안성만(Sung-Mahn Ahn)*, 김인환(In Hwan Kim)**, 최병구(Byounggu Choi)***, 조윤호(Yoonho Cho)****, 김은홍(Eunhong Kim)*****, 김명균(Myeong-Kyun Kim)******

초 록

협업필터링(collaborative filtering) 추천은 효과적인 추천을 위해 가장 널리 활용되는 기법 가운데 하나로 다양한 분야에서 널리 활용되고 있다. 협업필터링 추천과 관련하여 주요 이슈 가운데 하나는 왜 적용 도메인에 따라 추천 성과 간에 차이가 다르게 나타나는가이다. 이러한 추천 성과 간의 차이가 발생하는 원인에 대해 많은 연구들은 데이터의 특성에만 주목할 뿐 체계적인 설명을 제시하지 못하고 있는 것도 사실이다. 이러한 기존 연구의 문제점을 해결하기 위해 본 연구는 소셜네트워크의 구조적 측정 지표를 활용하여 추천 성과 간의 차이가 발생하는 원인을 보다 체계적으로 규명하고자 한다. 이를 위해 소셜네트워크의 구조적 측정지표와 협업필터링 추천 성과 간의 관계에 대한 가설을 수립하고 국내 H백화점의 거래데이터를 활용하여 이를 실증적으로 검증하였다. 검증 결과 밀도와 포괄성은 추천 성과에 긍정적인 영향을 미치는 반면 군집화계수는 부정적인 영향을 미치는 것을 파악하였다. 본 연구는 협업필터링 추천 성과를 이해할 수 있는 새로운 관점을 제시하였다. 또한 기업이 협업필터링 추천시스템을 도입하고자 할 때 그들의 의사결정에 도움을 줄 수 있는 가이드라인을 제시하였다는 점에서 그 의의가 있다.

ABSTRACT

Collaborative filtering (CF), one of the most successful recommendation techniques, has been used in a number of different applications such as recommending web pages, movies, music, articles and products. One of the critical issues in CF is why recommendation performances are different depending on application domains. However, prior literatures have focused on only data characteristics to explain the origin of the difference. Scant attentions have been paid to provide systematic explanation on the issue. To fill this research gap, this study attempts to systematically explain why recommendation performances are

본 연구는 2012년도 국민대학교 교내연구비 지원으로 수행됨.
* 국민대학교 경영대학 경영학부 부교수
** GS SHOP IT 연구개발팀
*** 교신저자, 국민대학교 경영대학 경영정보학부 부교수
**** 국민대학교 경영대학 경영정보학부 부교수
***** 국민대학교 경영대학 경영정보학부 교수
****** 국민대학교 경영대학 경영학부 교수
2012년 05월 03일 접수, 2012년 05월 16일 심사완료 후 2012년 05월 21일 게재확정.

different using structural indexes of social network. For this purpose, we developed hypotheses regarding the relationships between structural indexes of social network and recommendation performance of collaboration filtering, and empirically tested them. Results of this study showed that density and inconclusiveness positively affected recommendation performance while clustering coefficient negatively affected it. This study can be used as stepping stone for understanding collaborative filtering recommendation performance. Furthermore, it might be helpful for managers to decide whether they adopt recommendation systems.

키워드 : 협업필터링, 추천 성과, 소셜네트워크 분석, 밀도, 포괄성, 집중도, 군집화계수
Collaborative Filtering, Recommendation Performance, Social Network Analysis, Density, Inclusiveness, Clustering Coefficient

1. 서 론

협업필터링은 고객의 구매정보를 바탕으로 고객 간 구매 유사성을 분석함으로써 유사한 고객들을 파악하고 이들의 구매정보를 활용함으로써 추천대상 고객에게 적합하다고 판단되는 상품을 추천하는 기법을 말한다. Amazone.com과 CDNnow.com을 비롯한 수많은 기업들이 협업필터링 기법을 활용하여 고객에게 상품을 추천하고 있다[3, 23]. 그러나 협업필터링 기법을 기반으로 한 추천은 기업의 종류나 데이터 특성에 따라 추천 성과에 있어 많은 차이가 존재한다[21, 27]. 이와 같이 동일한 협업필터링 방식을 활용하더라도 추천 성과 간의 차이가 발생하는 원인으로는 데이터 희박성(data sparsity), 독특한 취향의 고객(gray sheep), 신규고객 추천의 문제, 신상품 추천 문제 등이 지적되고 있다[8, 27]. 그러나 이러한 원인들은 선호도와 관련된 정보가 불충분하기 때문에 고객 간 구매 선호도의 유사성 관계를 명확하게 추정할 수 없고 이로 인해 추천 성과간 차이가 존재한다는 일차원적인 설명은 할 수 있을지 모르나 실제 여러 요인에 근거하여 발생할 수 있는 추천 성과 간 차이를 명확하게 규명할 수 없다는 점이 한계로 지적되고 있다.

이러한 한계를 극복하기 위해 본 연구에서는 소셜네트워크 분석(social network analysis) 기법을 활용하고자 한다. 소셜네트워크 분석은 의사소통 집단 내 개체의 상호작용에 관심을 두고, 개체 간 연결 상태 및 연결 구조의 특성을 계량적으로 파악하여 시각적으로 표현하는 분석 기법이다[6]. 이 기법은 소셜네트워크 내의 정보 흐름과 구조적인 특성을 파악하고 이들이 갖는 의미를 해석 할 수 있는 장점이 있기 때문에 다양한 분야에 걸쳐 네트워크의 구조와 관계를 분석하는데 활용되어 왔다[22]. 협업필터링 기법에서는 고객 간 상품 선호도를 분석한 후 유사 고객을 연결하여 상품을 추천하는데 이 관계는 일종의 소셜네트워크라 할 수 있다[5, 7, 26]. 따라서 협업필터링 기법을 통해 구성되는 유사 고객 간 소셜네트워크의 구조적 특성이 협업필터링의 추천 성과와 어떠한 관계를 갖는지 파악함으로써 협업필터링 기반 추천 성과 간의 차이가 발생하는 원인을 보다 정밀하게

설명할 수 있을 것이다.

본 연구의 목적을 달성하기 위해 국내 유명 H 백화점의 거래 데이터를 기반으로 소셜네트워크를 구성하고 소셜네트워크의 구조적 측정지표가 추천 성과에 미치는 영향을 실증적으로 분석하고자 한다. 보다 구체적으로 기존 연구를 기반으로 소셜네트워크 측정지표에 따른 협업필터링 추천 성과에 대한 가설을 설정한 후 회귀분석을 실행하여 가설을 검증함으로써 추천 성능을 보다 정밀하게 예측할 수 있는 회귀모형을 도출하고자 한다. 이를 통해 i) 소셜네트워크 관점에서 협업필터링 추천 성과 간의 차이가 발생하는 원인에 대한 보다 명확한 이유를 규명하고 ii) 경영자로 하여금 협업필터링 활용의 효과성을 추천시스템 도입 이전에 예측 가능하도록 함으로써 추천시스템 도입과 관련된 의사결정에 도움을 줄 수 있을 것이다.

본 논문은 다음과 같이 구성되어 있다. 다음 장에서는 소셜네트워크 분석과 협업필터링 추천 관련 기존연구를 요약한다. 제 3장에서는 추천 성과와 소셜네트워크 측정지표 간의 관련성을 파악하고 이를 기반으로 가설을 제안한다. 제 4장에서는 제시된 연구가설을 검증하기 위한 연구방법을 설명하고 제 5장에서는 분석결과와 의미를 요약한다. 마지막으로 결론과 향후 연구 과제를 제안한다.

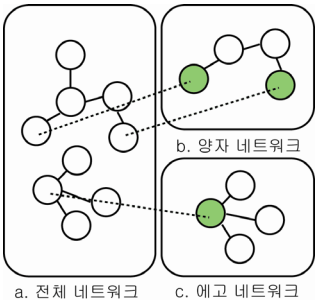
2. 문헌 연구

2.1 소셜네트워크 분석

소셜네트워크는 Barnes[9]에 의해 처음 사용된 용어로, 개인적인 인간관계가 확산되어 형성된 사람들 사이의 관계형태나 유형 혹은 구조를 말한다[6]. 소셜네트워크는 노드 (actors)와 관계 (ties)로 구성되는데, 노드는 소셜네트워크에 참여하는 각각의 참여자들을 의미하고 관계는 참여자들 사이에 가지는 사회적인 연결 관계를 의미한다. 소셜네트워크는 팀이나 조직, 산업등과 같은 다양한 사회적 구조의 행동을 설명하고 이해하는데 사용되어 왔다[22].

소셜네트워크는 유형에 따라 에고-네트워크(ego-centric network), 양자 네트워크(dyadic network), 전체 네트워크(total network)로 나눌 수 있다[6, 20, 32]. 에고-네트워크는 소셜네트워크의 한 구성원을 중심으로 그 구성원과 다른 구성원 간의 연결로 구성된 소셜네트워크를 말한다. 양자 네트워크는 소셜네트워크 내의 두 구성원을 중심으로 형성되는 연결로 구성되는 소셜네트워크를 말한다. 예를 들면 두 친구 사이의 관계나 한 중소기업과 주거래 대기업의 거래 관계를 들 수 있다. 전체 네트워크는 N명의 전체 행위자들로 구성된 보편적인 소셜네트워크를 말한다(<그림 1> 참조).

또한, 소셜네트워크는 관계의 방향성의 유무에 따라 방향성(directed) 네트워크와 비방향성(undirected) 네트워크로 표현된다. 방향성 네트워크는 시작점과 끝점이 존재하는 방향을 가진 관계를 포함하는 네트워크이다. 국가간 수입, 수출 관계를 나타내는 네트워크로 설명할 수 있다[31]. 비방향성 네트워크는 두 구성원 사이의 관계가 상호 동일한 네트워크이다. 구매 패턴의 유사성을 기반으로 형성된 고객 간 네트워크가 그 예가 될 수 있다.



〈그림 1〉 소셜네트워크의 유형

소셜네트워크의 구조는 응집성(cohesion), 구조적 등위성, 중심성 등 다양한 개념을 통해 파악 되어왔다. 예를 들면, 소셜네트워크의 구조 가운데 유사한 지위를 점하고 있는 행위자들 그룹화하고 그 그룹 간의 관계를 기준으로 소셜네트워크의 구조를 파악하기 위해 구조적 등위성 개념이 개발되었으며 소셜네트워크 구조가 중심에 위치하는 참여자에게 얼마나 집중되어 있는가를 파악하기 위해 소셜네트워크 중심성 개념이 개발되었다. 이러한 각 개념은 다양한 측정지표를 통해 측정할 수 있다. 예를 들면, 포괄성은(inclusiveness), 연결정도(degree), 밀도(density), 군집화 계수(clustering coefficient)와 같은 지표로 측정할 수 있으며[2, 6, 31], 중심성은 연결정도, 근접(closeness), 매개(betweeness) 중심성과 같은 지표로 측정할 수 있다[6, 16].

2.2 협업필터링 추천

협업필터링 추천은 가장 효과적인 추천방법으로 널리 알려져 있고 다양한 분야에서 활용되어 왔다[1, 3, 23]. 협업필터링 추천은 추천의 대상이 되는 고객과 구매 패턴이 유사한 고객의 구매정보를 취합하여 추천의 대상이 되는 고객이 구매할 가능성이 높을 것으로 예측되는 상품을 추천해 주는 방법이다.

협업필터링과 관련된 기존 연구는 <표 1>과 같이 크게 메모리 기반, 모델 기반, 하이브리드 연구로 분류할 수 있다[8]. 메모리 기반 협업필터링은 사용자 혹은 아이템 간 유사도를 측정하고 이를 기반으로 추천하는 방식으로 피어슨 상관계수(pearson correlation coefficient)나 코사인 벡터(cosine vector)가 널리 활용되고 있다[10, 15, 25, 29]. 모델 기반 협업필터링은 선형대수, 뉴럴 네트워크(neural network) 클러스터링 등을 기반으로 사전에 모델을 수립하여 추천하는 방식으로 클러스터링 협업필터링[10, 11], 잠재 시멘틱(latent semantic) 협업필터링[19, 24], 확률관계 모형 협업필터링[17] 등이 있다. 하이브리드 협업필터링은 기존 협업필터링 방식에 추가 정보를 제공할 수 있는 다른 추천기법을 결합함으로써 보다 향상된 추천 결과를 얻고자 하는 방식으로 내용기반 필터링과 협업필터링의 결합 또는 소셜네트워크 정보와 협업필터링의 결합방식 등이 있다[3, 12, 30]. 특히, 최근 들어 추천 성능을 향상시키기 위해 소셜네트워크의 정보와 협업필터링을 결합하는 연구가 활발하게 진행되고 있다[7]. 예를 들면, Liu and Lee[23]는 협업필터링에서 이웃의 형성을 소셜네트워크에서 직접 연결된 친구로만 제한함으로써 추천 성능이 향상됨을 주장하였다.

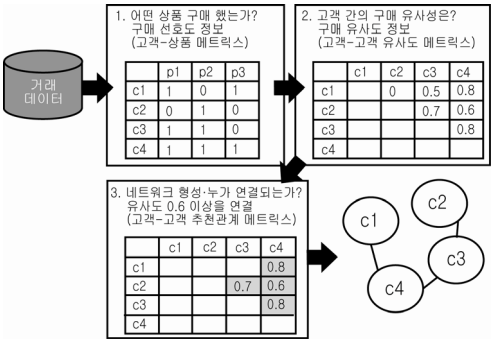
그러나 이러한 기존 협업필터링 관련 연구들은 협업필터링의 문제점을 개선하거나 추천의 성능을 향상시키는 방향에 초점을 맞춰 진행되어 왔을 뿐 추천 성과 간의 차이가 발생하는 원인에 대한 연구는 전혀 없었다[4, 7].

〈표 1〉 협업필터링 관련 기존 연구

범주	기법	연구
메모리 기반	아이템 기반 협업필터링 고객 기반 협업필터링	[10, 15, 25, 29]
모델 기반	클러스터링 협업필터링 잠재시멘틱 협업필터링 확률 관계 모형 협업필터링	[10, 11, 17, 19, 24, 27]
하이브리드	내용기반 협업필터링	[3, 12, 30]
	소셜네트워크 + 협업필터링	[5, 7, 23]

2.3 협업필터링과 소셜네트워크

협업필터링에서 가장 일반적인 알고리즘은 근접이웃 알고리즘이라 할 수 있다[8, 27]. 이는 피어슨 상관계수를 통해 근접 이웃을 찾고 해당 이웃에 대한 정보를 바탕으로 추천을 하는 알고리즘이다[8, 18]. 근접이웃 알고리즘은 고객 간 구매 연관성을 분석하고 이를 유사도로 묶어 상품 추천관계를 형성하게 되며 이를 네트워크로 표현하게 되면 일종의 소셜네트워크라 할 수 있다[5, 26]. <그림 2>는 근접이웃 알고리즘을 통한 소셜네트워크 형



〈그림 2〉 협업필터링에서의 소셜네트워크 형성

성과정의 한 예이다. 좀 더 구체적으로 표현하면 비방향성 이진(연결 형성 : 1, 연결 비형성 : 0) 관계인 소셜네트워크의 예라 할 수 있다.

3. 가설 개발

협업필터링 추천 성과의 차이를 발생시키는 원인 가운데 주요 원인으로는 데이터 희박성, 독특한 취향의 고객(gray sheep), 신규 고객 추천의 문제, 추천 범위(coverage), 우연성 추천 등이 지적되고 있다(<표 2> 참조). 본 연구에서는 협업필터링을 기반으로 형성된 소셜네트워크를 활용하기 때문에 협업필터링 추천 성과의 차이를 발생시키는 요인과 소셜네트워크의 측정지표 간의 관계를 바탕으로 가설을 수립하고자 한다.

3.1 데이터 희박성과 밀도

희박성은 상품과 고객의 수는 많은데 비해 고객이 구매한 데이터 자체가 너무 희박하여 고객 간 상품 추천관계를 형성하는데 문제가 있어 추천 성능이 저하될 수 있다는 것을 의미한다[3, 6]. 상품이 많아질수록 상품에 대한 고객의 평가정보나 구매정보를 통하여 수집되는 선호도 데이터가 존재하지 않는 상품의 수가 상대적으로 많아진다. 따라서 고객-상품 간의 행렬은 희박 행렬(sparse matrix)이 된다.[1] 이로 인해 유사한 이웃 집단을 탐색

1) 협업필터링 추천에서 과거 n명의 고객이 구매한 m개의 상품을 나타내는 거래를 m × n 고객-상품 행렬로 표현한다. 고객 간 구매 유사도는 고객-상품 간 행렬을 기반으로 피어슨 상관계수나 코사인 벡터를 통해 측정된다.

〈표 2〉 협업필터링에서 추천 성과의 영향 요인

영향요인	의미
데이터 희박성(sparsity)	상품 수에 비해 고객 선호도 데이터가 부족하여 고객 간 유사도 신뢰성이 떨어지고 결과적으로 추천 성과가 저하됨
독특한 취향의 고객 (gray sheep)	이질적이고 독특한 구매 행위를 갖는 고객의 경우 다른 고객과 유사도를 측정할 수 없어 추천 성과가 저하됨
신규고객 추천	신규고객의 경우 구매정보가 없어 다른 고객과 유사도를 측정할 수 없어 추천 성과가 저하됨
추천 범위(coverage)	추천 대상 고객에게 가능한 모든 후보 상품들을 추천할 수 없어 추천 성과가 저하됨
우연성(serendipity) 추천	고객이 과거에 구매하지 않았고 구매를 전혀 고려하지 않았지만 높은 흥미를 가질 수 있는 다른 종류의 상품 추천이 어려워 추천 성과가 저하됨

하는 과정에서 적은 수의 선호도 데이터만을 사용하게 되고 고객 간 유사도 측정의 신뢰성은 떨어지게 되며, 궁극적으로 추천 성과는 하락하게 된다[5, 8, 18].

밀도는 하나의 네트워크에 있어 참여자 간 가능한 모든 관계의 수 가운데 실제로 맺어진 관계의 비율로 정의된다[2, 6]. 구매 패턴의 유사성을 기반으로 형성된 고객 간 네트워크에서, 밀도는 고객 사이에 얼마나 유사한 구매패턴 관계를 가지고 있는가로 해석될 수 있다. 즉, 소셜네트워크에 존재하는 관계가 많을수록 고객 간 유사한 구매 패턴 관계가 많다는 것을 의미한다.

희박성 관점에서 밀도의 증가는 희박성의 감소를 의미한다. 왜냐하면 밀도의 증가는 고객 간 관계의 증가를 의미하고, 이는 적어도 두 고객 간 구매 선호도 데이터가 유사성을 측정할 수 있는 수준 이상이라는 것을 의미한다. 따라서 밀도가 증가하면 증가할수록 행렬의 빈 셀들은 감소하게 되며 이는 희박성의 감소를 의미한다. 이를 정리하면, 소셜네트워크에서 밀도의 증가는 희박성의 감소를 의미하며, 이는 결과적으로 추천 성과의 향상을 의미한다. 따라서 본 연구에서는 다음과 같은 가설을 제안한다.

H1 : 밀도가 증가하면 추천 성능은 향상 될 것이다.

3.2 독특한 취향의 고객(gray sheep) 및 신규고객 추천과 포괄성

독특한 취향의 고객(gray sheep)은 다른 고객들과 전혀 다른 자신만의 구매패턴을 보이는 고객을 의미한다. 이러한 고객들은 다른 고객과 동일하게 구매한 제품이 없어 구매 유사성을 도출하기가 매우 어렵다. 이러한 문제는 신규고객의 경우에도 마찬가지다. 신규 고객의 경우 기존 구매 이력이 없기 때문에 구매에 관한 선호도 데이터를 파악할 수 없고 결과적으로 구매 도출하기가 불가능하다. 따라서 독특한 취향의 고객과 신규 고객이 증가하면 할수록 추천 성과는 떨어지게 된다[5, 8].

포괄성은 하나의 네트워크에 포함된 참여

자의 총 수에서 연결되지(isolates) 않은 참여자의 수를 뺀 연결된 참여자들의 비율로 정의된다[6]. 본 연구에서 구성되는 소셜네트워크는 고객의 구매 이력에 기반한 구매 패턴의 유사성을 관계로 정의하기 때문에 독특한 취향의 고객과 신규고객은 다른 고객들과 관계를 형성하지 못하는 고립된 고객으로 간주할 수 있다.

독특한 취향의 고객과 신규고객이라는 관점에서 보면, 포괄성은 증가는 독특한 취향의 고객과 신규고객의 감소를 의미한다. 왜냐하면, 포괄성의 증가는 고립된 고객들의 수가 감소함을 의미하고 이는 유사한 구매 패턴의 관계를 형성하는 고객의 비율이 증가함을 의미한다. 이를 종합하면, 포괄성이 증가는 독특한 취향의 고객 또는 신규고객과 같은 고립된 고객 수의 감소를 의미하며, 이는 결과적으로 협업필터링 추천 성과의 향상을 의미한다. 이를 근거로 다음과 같은 가설을 제안한다.

H2 : 포괄성이 증가하면 추천 성능이 향상될 것이다.

3.3 추천 범위 및 우연성 추천과 군집화 계수

추천 범위(coverage)는 전체 상품 중 협업필터링 추천을 통해 추천 대상 고객에게 추천해 줄 수 있는 상품의 비율을 말한다. 추천 범위가 넓을수록 추천 대상 고객의 상품 탐색 기회는 증가하게 되고[18], 결과적으로 고객의 입맛에 맞는 추천 기회를 높여 추천 성과의 향상을 가져올 수 있다. 우연성(serendipity) 추천은 추천 대상 고객에게 과거에 구매하지 않았고 전혀 고려하지 않았지만 높은 흥미를 가질 수 있는 다른 종류의 상품을 추천하는 것을 의미한다[18]. 이러한 우연성 추천은 추천의 정확성보다는 추천의 만족도를 높이는 역할을 한다.

군집화계수는 네트워크 내의 3명의 참여자 a, b, c가 있고 a와 b, a와 c사이에 관계가 있을 때 b와 c가 관계를 가질 가능성을 말한다. 예를 들면, 친구관계 네트워크에 있어, 한 사람의 친구들이 서로 친구가 될 가능성을 의미한다. 본 연구에서 군집화계수는 추천대상 고객의 이웃들 간의 관계가 형성된 정도로 파악할 수 있다. 군집화 계수가 높다는 것은 이웃들 간 관계가 많이 형성되어 있다는 것을 의미하며 이는 그들 간 구매 패턴이 유사하다는 것을 의미한다.

추천범위 및 우연성 추천의 관점에서 보면, 군집화 계수가 낮다는 것은 이웃들의 구매패턴이 다양하다는 것을 의미한다. 이러한 구매패턴의 다양성은 추천 범위를 증대시킬 뿐 아니라 우연성 추천에 대한 범위도 증대시킨다. 따라서 군집화 계수가 낮아질수록 다수의 상품을 평가할 수 있는 기회와 새로운 상품을 접할 기회를 향상시킴으로써 추천의 정확성 및 만족도를 향상시킬 가능성이 높아진다. 이를 근거로 본 연구에서는 다음과 같은 가설을 제안한다.

H3 : 군집화 계수가 낮을수록 추천 성능은 향상될 것이다.

3.4 추천 범위와 집중도

집중도(degree centralization)는 네트워크에

〈표 3〉 가설개발의 주요 논거

가설	주요 논거
H1 : 밀도가 증가하면 추천 성능은 향상될 것이다.	밀도의 증가는 희박성의 감소를 의미하며, 이는 결과적으로 추천 성과의 향상을 가져올 것임
H2 : 포괄성이 증가하면 추천 성능이 향상될 것이다.	포괄성이 증가는 독특한 취향의 고객 또는 신규고객과 같은 고립된 고객 수의 감소를 의미하며, 이는 결과적으로 추천 성과의 향상을 가져올 것임
H3 : 군집화 계수가 낮을수록 추천 성능은 향상될 것이다.	군집화 계수가 낮아질수록 추천 범위 넓어지고 우연성 추천의 기회를 많아져 추천의 정확성 및 만족도를 향상시킬 가능성이 높아질 것임
H4 : 집중도가 낮을수록 추천 성능은 향상될 것이다.	집중도가 낮을수록 추천 범위가 넓어질 것이며 이는 추천 성능을 가져올 것임

서 중심성이 높은 참여자에게 얼마나 네트워크 관계가 집중되어 있는지를 의미한다[6, 16].[2] 본 연구에서 집중도는 네트워크 형성에 있어 얼마나 많은 고객 의견이 반영되었는가로 파악할 수 있다. 본 연구에서 추천 성과는 네트워크 형성에 참여한 개개인의 추천 성과의 평균이다. 따라서 각 참여자의 추천 성과가 좋아야 전체 네트워크의 추천 성과도 좋아질 것이다.

추천 범위 관점에서 파악해 보면, 집중도가 높다는 것은 핵심고객에 대한 의존도가 높다는 것을 의미한다. 네트워크의 집중도가 높아 핵심고객에 의존적이게 되면 소수 핵심고객의 추천 성능은 많은 변두리 고객들의 의견을 취합할 수 있어 향상되는 반면 다수의 변두리 고객들은 소수 핵심고객의 의견을 기반으로 추천을 형성하기 때문에 추천의 범위가 매우 제한적이기 된다. 앞서 언급한 바와 같이 추천 범위의 감소는 추천 성능의 저하를 가져온다. 따라서 본 연구에서는 다음과 같은 가설을 제안한다.

H4 : 집중도가 낮을수록 추천 성능은 향상될 것이다.

협업필터링에서 추천 성능에 영향을 주는 요인과 소셜네트워크 측정지표 간의 이상의 논의를 요약하면 다음 <표 3>과 같다. 다양한 기존 문헌에 따르면 추천 대상의 직접적인 이웃을 고려하여 형성된 유사한 구매패턴의 고객 간 관계는 하나의 소셜네트워크라 할 수 있다[5, 7, 23]. 따라서 이렇게 형성된 소셜네트워크 구조는 추천 대상의 직접적인 이웃이 어떻게 결정되는가에 영향을 받게 된다. 이를 근거로 이웃 형성에 영향을 주는 요인인 데이터 희박성, 독특한 취향의 고객, 신규고객 추천, 추천 범위, 우연성 추천은 소셜네트워크 구조에 영향을 미치는 것으로 판단할

2) 집중도는 일반적으로 연결정도 집중도, 매개 집중도, 근접 집중도가 있다. 다만, 협업필터링에서는 추천 대상의 직접적인 이웃만을 고려하여 추천을 형성기 때문에 본 연구에서는 직접적인 관계의 집중도를 나타내는 연결정도 집중도만을 고려한다. 연결정도 집중도는 0에서1 사이의 값을 갖는데 네트워크 구조가 관계를 많이 갖는 참여자에게 집중될수록 값은 1에 가깝게 된다. 극단적으로 스타형 구조 네트워크의 집중도 값은 1이며 원형 구조 네트워크의 집중도 값은 0이다.

수 있으며 이는 소셜네트워크 측정지표를 통해 파악할 수 있다.

4. 연구 방법

4.1 표본 및 자료수집

본 연구에서 제안한 가설 검증을 위해 다양한 소셜네트워크 구조를 가지고 있는 국내 유명 H 백화점의 1년간(2000년 5월 1일부터 2001년 4월 30일까지) 거래데이터를 표본으로 자료를 수집하였다. 전체 1,660,814건의 거래 내역(50,000명의 고객이 구매한 4,038개의 상품)을 기반으로 4단계 자료수집 절차를 소셜네트워크를 형성하였다(<표 4> 참조).

〈표 4〉 자료수집 절차단계와 단계별 샘플 수

샘플링 단계	샘플링 방법	샘플의 수
1단계	월별 분할	12
2단계	2개월씩 조합	66(12×11/2)
3단계	무작위 고객의 수 100, 200, 300으로 샘플링	198(66×3)
4단계	소셜네트워크 형성 시, 연결 기준인 유사도 임계치(0.1, 0.2) 활용	396(198×2)

첫째, 백화점의 거래데이터의 경우 월별로 거래량이 다르고 포함되는 고객과 상품의 종류가 다르기 때문에 전체 표본을 12달의 거래데이터로 분할하였다. 둘째, 샘플의 확보를 위해 2개월을 조합하여 66($_{12}C_2$)개의 샘플을 도출하였다. 셋째, 각 샘플에 대해 포함되어 있는 고객의 수가 100, 200, 300명인 샘플을 추출하여 198(66×3)개를 도출하였다. 이때, 최대 크기는 300명으로 제한하였다. 이는 고객의 수가 지나치게 많아지는 경우 소셜네트워크 측정지표와 추천 성과 측정을 위한 연산시간이 기하급수적으로 증가하기 때문이다. 마지막으로 소셜네트워크 관계 형성에 있어 유사도의 임계치 수준을 0.1과 0.2의 두 수준을 고려하였다. 이는 임계치 수준이 0.3 이상일 경우 소셜네트워크 관계의 희박성이 증가되어 구조 자체가 의미를 잃기 때문이다. 결론적으로 총 396(198×2)개의 데이터 셋이 가설 검증을 위해 활용되었다.

4.2 소셜네트워크 형성과정

유사한 구매패턴 관계를 가지는 고객 간 소셜네트워크를 구성하기 위해 먼저 고객 구매패턴 간 유사도를 분석할 필요가 있다. 고객의 구매패턴 분석을 위해 거래데이터로부터 고객이 어떤 상품을 구매했는가를 파악하고 이를 기반으로 식 (1)과 같은 고객-상품 행렬을 구성한다. 이때 고객이 단일상품을 여러 번 구매한다 하여도 한 번의 구매로 간주한다.

$$p_{ij} = \begin{cases} 1 : \text{고객 } i\text{가 상품 } j\text{를 구매} \\ 0 : \text{고객 } i\text{가 상품 } j\text{를 비구매} \end{cases} \quad (1)$$

고객-상품 행렬이 완성되면 자카드(Jaccard) 유사도를 통해 고객 간 유사도를 계산한다(식 (2) 참조).[3] 자카드 유사도는 0과 1사의 값을 갖게 되며 1이면 두 고객 사이의 구매

3) 이진 데이터로 이루어진 집합 간 유사도를 계산하는 경우 자카드 유사도가 더 용이하고 직관적이다[11].

가 일치함을 0이면 전혀 다름을 의미한다.

$$sim(a,\ b) = J = (a,b) = \frac{M_{11}}{M_{10} + m_{01} + M_{11}} \quad (2)$$

M_{11} : 고객 a, b가 모두 구매한 상품
M_{01} : 고객 b만 구매한 경우
M_{10} : 고객 a만 구매한 경우
M_{00} : 고객 a, b 모두 구매하지 않은 경우

자카드 유사도를 기반으로 유사한 구매패턴의 고객이 파악되면 이들을 연결함으로써 소셜네트워크를 형성할 수 있다. 이때, 유사도 임계치 ρ를 기준으로 임계치 이상의 고객을 연결함으로써 소셜네트워크를 형성할 수 있다. 앞서 언급한 바와 같이 0.3 이상의 임계치를 활용할 경우 고객 간 관계의 희박성 매우 높아지지 때문에 본 연구에서 0.1과 0.2 두 개의 임계치를 활용한다.

4.3 변수의 측정

본 연구의 독립변수인 밀도, 포괄성, 연결정도 집중도, 군집화계수는 기존 연구를 준용하여 측정하였다(<부록 1> 참조). 측정의 편의를 위해 툴인 Ucinet v6와 Netminer 3.0을 활용하였다.

본 연구의 종속변수인 협업필터링 추천 성과를 측정하기 위해 가장 널리 활용되는 메모리 기반 알고리즘 [27]을 적용하여 추천시스템을 구축한 후, 구축된 추천시스템을 396개의 데이터 셋 각각에 적용하여 추천 정확도를 측정하였다. 각 데이터 셋은 2개월간의 거래 데이터를 포함하고 있기 때문에 거래일자가 빠른 순으로 40일 간의 거래 데이터를 훈련(training) 데이터로 나머지 20일의 거래 데이터를 검정(test) 데이터로 분할한 후, 훈련 데이터에 메모리기반 추천 알고리즘을 적용하여 각 고객들에게 추천할 상품을 선정하고 이를 검정 데이터의 실제 구매 상품 과 비교하여 추천 성과를 측정하였다.

추천 성과 측정을 위한 지표로는 재현율(*recall*)[4]과 정확률[5](*precision*)을 동일한 가중치로 결합한 F1-척도(*measure*)를 사용하였다[5, 18].

$$F1-\text{척도}(measure) = \frac{2 \times \text{재현율}(recall) \times \text{정확률}(precision)}{(\text{재현율}(recall) \times \text{정확률}(precision))} \quad (3)$$

4.4 분석절차

가설 검증을 다중회귀 분석을 실시하였다. 또한 최종 회귀모형은 단계선택법(stepwise)을 적용하여 도출하였다. <표 5>는 변수에 대한 기초통계량과 변수 간 상관관계를 나타내고 있다. 상관관계 분석결과 밀도와 포괄성(0.892) 간의 매우 높은 상관관계가 있음이

4) 재현율은 추천 대상고객이 검정 데이터 집합에서 구매한 상품 중에서 협업필터링 알고리즘에 의해 추천된 상품의 비율로 정의되며 다음 식 (f1)과 같이 나타낼 수 있다(Sarwar et al., 2001).

$$\text{재현율}(recall) = \frac{|\text{검정 데이터 집합에서 구매한 상품} \cap \text{추천한 상품}|}{\text{검정 데이터 집합에서 구매한 상품}} \quad (f1)$$

5) 재현율은 추천 대상고객이 검정 데이터 집합에서 구매한 상품 중에서 협업필터링 알고리즘에 의해 추천된 상품의 비율로 정의되며 다음 식 (f1)과 같이 나타낼 수 있다(Sarwar et al., 2001).

$$\text{정확률}(\text{Pr}ecision) = \frac{|\text{검정 데이터 집합에서 구매한 상품} \cap \text{추천한 상품}|}{\text{추천한 상품}} \quad (f2)$$

파악되었다. 따라서 분산확대인자(VIF : variance inflation factor)를 통해 독립변수 간 다중공선성(multicollinearity)이 있는지를 파악하였다. 분석결과 모든 변수의 VIF 값이(밀도 = 5.978, 포괄성 = 5.821, 군집화계수 = 1.220, 집중도 = 1.305) 10 이하로 다중공선성이 크게 문제가 되지 않는 것으로 나타났다.

〈표 5〉 기초통계량 및 상관관계 분석

변수	평균 (표준편차)	밀도	포괄성	군집화계수	집중도	F1-척도
밀도	.082 (.065)	1				
포괄성	.702 (.235)	.892**	1			
군집화계수	.568 (.079)	.165**	-0.030	1		
집중도	.198 (.102)	.466**	.472**	.063	1	
F1-척도	.016 (.012)	.825**	.887**	-0.055	.404**	1

** 0.01수준에서 유의.

5. 연구결과 및 논의

5.1 분석결과

다음 <표 6>은 회귀분석 결과를 나타내고 있다. R^2 값에서 알 수 있듯이 79.7%의 변동성이 4개의 변수에 의해 설명됨을 알 수 있다.

회귀분석 결과 밀도는 추천 성과에 유의한 영향을 미치는 것으로 파악되었다($p < 0.001$). 따라서 가설 1은 채택된다. 포괄성 역시 추천 성과에 유의한 영향을 미치는 것으로 파악되었다($p < 0.001$). 또한 예측한 바와 같이 군집화 계수는 추천 성과에 유의한 음의 영향을 미치는 것으로 파악되었다($p < 0.01$). 따라서 가설 2와 3은 채택된다. 그러나 집중도는 추천 성과에 유의한 영향을 미치지 않는 것으로 나타났다($p > 0.1$). 따라서 가설 4는 기각한다.

〈표 6〉 분석결과

변수	표준화계수 (베타)	t-값	유의확률	가설검증결과
상수항		-1.588	.113	
밀도	.237	4.254	.000	H1 채택
포괄성	.685	12.350	.000	H2 채택
군집화계수	-.072	-2.883	.004	H3 채택
집중도	-.025	-.976	.330	H4 기각

R^2 = 0.798; F-값 = 385.32($p < 0.001$).

5.2 논의

본 연구 결과 가설 1은 채택되었다. 즉 밀도가 증가는 데이터 희박성이 감소를 의미하며 이는 협업필터링 추천 성과의 향상을 의미한다고 할 수 있다. 따라서 고객 간 구매패턴의 유사성을 기반으로 형성된 소셜네트워크에서 참여자간 관계가 많으면 많을수록 협업필터링을 통한 추천 성과가 보다 향상될 것이다. 연구 가설 2 역시 채택되었다. 즉 포괄성이 증가는 독특한 취향의 고객 또는 신규고객과 같은 고립된 고객 수가 감소를 의미하며 이는 협업필터링 추천 성과의 향상을 의미한다고 할 수 있다. 따라서 고객 간 구매패턴 유사성을 기반으로 형성된 소셜네트워크에서 다른 참여자와 연결되지 않은 고립된

참여자가 적은 경우 협업필터링 알고리즘을 적용하여 추천할 경우 이의 성과는 매우 높을 것이다. 군집화계수와 추천 성과 간의 관계를 파악한 연구 가설 3 역시 채택되었다. 이는 군집화계수가 낮아질수록 다수의 상품을 평가할 수 있는 기회와 새로운 상품을 접할 기회가 향상되고 결과적으로 추천의 정확성 및 만족도가 높아질 것이기 때문이다. 따라서 소셜네트워크가 서로 유사한 구매패턴으로 연결되어 있으나 추천 대상의 이웃 고객 간의 구매는 유사하지 않은 경우 추천 성과는 향상될 것이다.

본 연구에서는 네트워크의 집중도가 높아질수록 핵심고객에 대한 의존성이 높아지게 되어 소수 핵심 고객의 추천 성능은 향상되는 반면 다수의 변두리 고객들의 추천 성능이 저하될 것으로 예측하였다. 그러나 연구결과 집중도는 추천 성과에 유의미한 영향을 미치지 않는 것으로 파악되었다. 이러한 흥미로운 결과는 실제 현실에서 발생하는 소셜네트워크의 구조가 원형이나 스타형과 같은 극단적인 형태가 아니라 보다 복합적인 구조일 가능성이 높기 때문인 것으로 설명할 수 있다. 즉, 실제 소셜네트워크는 스타형이나 원형과 같은 극단적인 형태를 가지는 경우가 드물고 집중도가 0인 경우에도 각 고객이 가지는 연결 관계의 수에 따라 추천 범위가 달라질 수 있기 때문에 집중도만으로 추천 성과를 평가하기에는 한계가 있기 때문일 것으로 추측된다.

6. 결 론

본 연구는 협업필터링 추천 성과에 영향을 미치는 소셜네트워크의 구조적 측정지표에 대해 고찰하였다. 기존 협업필터링에서 연구되어 온 추천 성과에 영향을 미치는 요인(데이터 희박성, 독특한 취향의 고객, 신규고객 추천의 문제, 추천 범위, 우연성 추천)을 기반으로 그에 대응하는 소셜네트워크의 구조적 측정지표들(밀도, 포괄성, 군집화계수, 집중도)을 파악하였다. 나아가 소셜네트워크 측정지표와 추천 성과 간의 관계에 대한 가설을 도출하였으며 이를 회귀모형을 통해 검증하였다. 분석결과 밀도와 포괄성은 추천 성과에 긍정적인 영향을 미치고 군집화계수는 부정적인 영향을 미치는 것으로 나타났다. 이를 통해 고객 간 구매 선호도 유사성을 기반으로 형성한 소셜네트워크에서는 고객 간 관계가 많을수록, 고립된 고객들이 적을수록, 추천 대상 고객의 이웃 간 구매가 이질적일수록 추천 성과는 향상됨을 알 수 있었다. 본 연구는 협업필터링 추천 성과를 소셜네트워크 구조적인 관점에서 이해함으로써 추천 성과 간의 차이가 발생하는 원인을 이론적으로 규명하고자 하는 초석이 되는 연구라는 점에서 그 의의가 있을 것이다. 또한 실무적인 관점에서 소셜네트워크의 구조를 파악함으로써 협업필터링의 추천 성과를 예측할 수 있는 가이드라인을 제시하였다는 점에서 그 의의가 있을 것이다. 즉, 도출된 회귀식을 기반으로 기업의 추천시스템 도입 시 성과를 사전에 예측해 볼 수 있기 때문에, 기업 입장에서 막대한 비용이 요구되는 추천시스템의 도입 여부를 결정할 수 있도록 도와줄 수 있을 것으로 판단된다.

본 연구는 다음과 같은 한계점이 있다. 첫째, H 백화점의 거래데이터만을 이용하여 분

석을 진행하여 해당 거래데이터에 종속적인 결과가 도출됐을 가능성이 있다. 따라서 향후 연구에서는 이러한 점을 보완하여 다양한 기업의 거래데이터를 활용하는 연구가 필요하다. 둘째, 비록 VIF 검정결과 다중공선성에 큰 문제가 없는 것으로 파악되었으나 밀도와 포괄성의 VIF 값이 상대적으로 높은 것으로 나타났다. 따라서 이에 대한 면밀한 검토가 필요할 것이다.

참 고 문 헌

[1] 강주영, 김현구, 박상언, “협업필터링 기반의 휴대폰 무선 서비스 추천을 위한 기지국 군집분석과 검증”, 전자거래학회지, 제15권, 제2호, pp. 1-18, 2010.

[2] 김용학, 사회연결망 분석, 박영사, 2003.

[3] 김재경, 조윤호, 김승태, 김혜경, “모바일 전자상거래 환경에 적합한 개인화된 추천시스템”, 경영정보학연구, 제15권, 제3호, pp. 223-241, 2005.

[4] 김형도, “일관성 기반의 신뢰도 정의에 의한 협업필터링”, 전자거래학회지, 제15권, 제2호, pp. 1-11, 2009.

[5] 박종학, 조윤호, 김재경, “사회연결망 : 신규고객 추천문제의 새로운 접근법”, 지능정보연구, 제15권, 제1호, pp. 123-139, 2009.

[6] 손동원, 소셜네트워크 분석, 경문사, 2002.

[7] 조윤호, 방정혜, “신상품 추천을 위한 사회연결망분석의 활용”, 지능정보연구, 제15권, 제4호, pp. 183-200, 2009.

[8] Adomavicious, 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, pp. 734-749, 2005.

[9] Barnes, J. A., “Class and Committees in a Norwegian Island Parish,” Human Relations, Vol. 7, pp. 39-58, 1954.

[10] Breese, J. S., Heckerman, D., and Kadie, C., “Empirical Analysis of Predictive Algorithms for Collaborative Filtering,” In Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence (UAI-98), San Francisco, California, pp. 43-52, 1998.

[11] Chee, S. H. S., Han, J., and Wang, K., “RecTree : An Efficient Collaborative Filtering Method,” Lecture Notes in Computer Science, Vol. 2114, pp. 141-151, 2001.

[12] 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. 3, pp. 234-246, 2004.

[13] Choi, S., Cha, S., and Tappert, C. C., “A Survey of Binary Similarity and Distance Measures,” Journal of Systemics, Cybernetics and Informatics, Vol. 8, No. 1, pp. 43-48, 2010.

[14] De Amorim, S., Barthelemy, J. P., and Ribeiro, C., “Clustering and Clique Parti-

tioning : Simulated Annealing and Tabu Search Approaches," Journal of Classification, Vol. 9 , pp. 17-41, 1992.

[15] Delgado, J. and Ishii, N., "Memory-based Weighted-majority Prediction for Recommender Systems," In ACM SIGIR'99 Workshop on Recommender Systems : Algorithms and Evaluation, 1999.

[16] Freeman, L., "Centrality in Social Networks : Conceptual Clarification," Social Networks, No. 1, pp. 215-239, 1979.

[17] Getoor, L. and Sahami, M., "Using Probabilistic Relational Models for Collaborative Filtering," In Workshop on Web Usage Analysis and User Profiling (WEBKDD'99), 1999.

[18] Herlocker, J. L., Konstan, J. A., Terveen, L. G., and Riedl, J. T., "Evaluating Collaborative Filtering Recommender Systems," ACM Transactions on Information Systems, Vol. 22, No. 1, pp. 5-53, 2004.

[19] Hofmann, T., "Latent Semantic Models for Collaborative Filtering," ACM Transactions on Information Systems, Vol. 22, No. 1, pp. 89-115, 2004.

[20] Human, S. E. and Provan, K. G., "Legitimacy Building in the Evolution of Small-firm Multilateral Networks : A Comparative Study of Success and Demise," Administrative Science Quarterly, Vol. 45, No. 2, pp. 327-365, 2000.

[21] Huang, Z. and Zeng, D., "Why Does Collaborative Filtering Work? Recommendation Model Validation and Selection by Analyzing Random Bipartite Graphs," the Fifteenth Annual Workshop on Information Technologies and Systems, 2005.

[22] Kukkonen, H. O., Lyytinen, K., and Yoo, Y. J., "Social Networks and Information Systems : Ongoing and Future Research Streams," Journal of the Association for Information Systems, Vol. 11, Special Issue, pp. 61-68, 2010.

[23] Liu, F. and Lee, H. J., "Use of Social Network Information to Enhance Collaborative Filtering Performance," Expert Systems with Applications, Vol. 37, pp. 4772-4778, 2010.

[24] Marlin, B., "Modeling User Rating Profiles for Collaborative Filtering," In Proceedings of the 17th Annual Conference on Neural Information Processing Systems (NIPS '03), 2003.

[25] Resnick, P., Iakovou, N., Sushak, M., Bergstrom, P., and Riedl, J., "GroupLens : An Open Architecture for Collaborative Filtering of Netnews," In Proceedings of the 1994 Computer Supported Cooperative Work Conference, pp. 174-186, 1994.

[26] 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, pp. 115-120, 2006.

[27] Sarwar, B., Karypis, G., Konstan, J. A., and Riedl, J., "Analysis of Recommenda-

tion Algorithms for E-commerce," Proceedings of ACM E-commerce 2000 conference, pp. 158-167, 2000.

[28] Schank, T. and Wagner, D., "Approximating Clustering Coefficient and Transitivity," Journal of Graph Algorithms and Applications, Vol. 9, No. 2, pp. 265-275, 2005.

[29] Shardanand, U. and Maes, P., "Social Information Filtering : Algorithms for Automating 'Word of Mouth'," In Proceedings of the Conference on Human Factors in Computing Systems, pp. 210-217, 1995.

[30] Soboroff, I. and Nicholas, C., "Combining Content and Collaboration in Text Filtering," In IJCAI '99 Workshop : Machine Learning for Information Filtering, pp. 86-91, 1999.

[31] Wasserman, S. and Faust, K. Social Network Analysis : Methods and Application, New York : Cambridge University Press, 1994.

[32] Weare, C., Loges, W. E., and Oztas, N., "Email Effects on the Structure of Local Associations : A Social Network Analysis," Social Science Quarterly, Vol. 88, No. 1, pp. 222-243, 2007.

〈부록 1〉 독립변수의 측정

1. 포괄성과 밀도

포괄성과 밀도를 산출하기 위해 각각 다음 식 (a-1)과 식 (a-2)를 활용하였다[2, 6, 31]. 식 (a-2)에서 n은 네트워크 전체 참여자의 수이고 k는 실제 연결된 관계의 수이다.

$$\text{포괄성비율} = \frac{(\text{연결된 점의 수})}{(\text{네트워크 전체의 수})} \quad \text{(a-1)}$$

$$\text{밀도}(density) = \frac{k}{n(n-1)/2} \quad \text{(a-2)}$$

2. 군집화계수의 측정

네트워크 내의 3명의 참여자 a, b, c를 가정하자. 이때 a와 b 그리고 a와 c사이에 관계가 있는 경우를 트리플(triple)이라고 하며 a, b, c 사이에 모든 관계가 연결되어 있는 경우를 트라이앵글(triangle)이라고 한다[28]. 군집화계수를 구하기 위해 먼저 네트워크에 있는 각 노드의 트라이앵글 수를 트리플 수로 나누어 계산한다. 예를 들면, a의 군집화 계수는 식 (a-3)과 같다. 다음으로 개별 참여자의 군집화계수 값을 평균한 값이 네트워크 수준의 군집화계수가 된다. 즉, 전체 네트워크의 군집화계수는 식 (a-4)를 활용하면 도출 할 수 있다[26].

$$c(a) = \frac{a\text{의 } triangle}{a\text{의 } tiple} \quad \text{(a-3)}$$

$$CC = \frac{1}{|V|}\sum_{a \in V} c(a) \quad \text{(a-4)}$$

주의 : 이때 V는 차수가 2 이상인 노드의 집합이다.

3. 연결정도 집중도의 측정

연결정도 집중도를 측정하기 위해서는 연결정도 중심성을 먼저 파악할 필요가 있다. 연결정도 중심성은 한 참여자에게 직접 관계를 맺는 노드들의 수를 의미한다. 따라서, 어떤 참여자가 연결정도 중심성이 높다는 것은 해당 참여자와 직접적인 관계를 맺는 다른 참여자가 많다는 것을 의미한다. 예시를 위해 참여자 Pi와 Pk를 가정하자. 이때, Pi와 Pk가 직접적으로 연결되면 1 값을 아니면 0 값을 갖는 함수를 $a(P_i,\ P_k)$로 정의하면, 참여자 Pk의 연결정도 중심성은 다음 식 (a-5)

와 같이 나타낼 수 있다[14].

$$C_D(P_k) = \sum_{i=1}^{n} a(P_i,\ P_k) \tag{a-5}$$

연결정도 집중도는 연결정도 중심성 높은 참여자에 얼마나 많은 네트워크 관계가 집중되어 있는지를 의미하며 식 (a-6)과 같은 방법을 통해 계산될 수 있다.

$$NC_D = \frac{\sum_{i=1}^{n}([C_D(P^*) - C_D(P_l)])}{\max\sum_{i=1}^{n}[C_D(P^*) - C_D(P_i)]} = \frac{\sum_{i=1}^{n}[C_D(P^*) - C_D(P_i)]}{[(n-1)(n-2)]} \tag{a-6}$$

이때, $C_D(P^*)$은 네트워크에서 나올 수 있는 가장 높은 연결정도 중심성 값을 의미한다. 따라서 분자의 $C_D(P^*)$는 현재 네트워크에서 가장 높은 연결정도 중심성을 의미하고 분모의 $C_D(P^*)$는 이론적으로 가능한 연결정도 중심성의 최대치를 의미한다. 따라서 분자의 경우 이론적으로 가능한 최대 연결정도 중심성은 스타형태의 네트워크 구조에서 발생하게 되며 그 최대치는 (n-1)(n-2)로 나타낼 수 있다. 이때 $C_D(P_i)$는 참여자 i의 연결정도 중심성 값을 의미한다[6, 16].

저 자 소 개

안성만 (E-mail : sahn@kookmin.ac.kr)
1980년~1984년 서울대학교 경영학과 (학사)
1984년~1986년 KAIST 경영과학과 (석사)
1992년~1998년 George Mason University (경영학박사)
2000년~2001년 동국대학교 정보관리학과 조교수
2001년~현재 국민대학교 경영대학 부교수
관심분야 통계적방법론, 데이터마이닝, 확률분포추정6

김인환 (E-mail : ih.kim@gsshop.com)
2002년~2009년 국민대학교 경영대학 e-비즈니스 학과 (학사)
2009년~2011년 국민대학교 e-비즈니스 대학원 (석사)
2011년~현재 GS 홈쇼핑
관심분야 사회연결망 분석, 데이터마이닝, 추천방법론, 고객행동분석, 대용량 데이터 분석

최병구 (E-mail : h2choi@kookmin.ac.kr)
1990년~1994년 고려대학교 정경대학 통계학과 (학사)
1994년~1996년 KAIST 경영대학원 경영공학전공 (석사)
1996년~2002년 KAIST 경영대학원 경영공학전공 (박사)
2002년~2003년 University of Minnesota Carlson School of Management (방문연구원)
2004년~2008년 University of Sydney, School of Information Technologies 조교수
2008년~2009년 국민대학교 경영대학 조교수 (e-비즈니스전공)
2010년~현재 국민대학교 경영대학 부교수 (전자상거래전공)
관심분야 지식경영, 인터넷비지니스, 정보시스템 가치평가

조윤호 (E-mail : www4u@kookmin.ac.kr)
1984년~1988년 서울대학교 계산통계학과 (학사)
1988년~1993년 LC전자 (주임연구원)
1994년~1996년 KAIST 경영대학원 경영공학전공 (석사)
1996년~2002년 KAIST 경영대학원 경영공학전공 (박사)
2004년~2006년 국민대학교 경영대학 조교수 (e-비즈니스전공)
2006년~현재 국민대학교 경영대학 부교수 (전자상거래전공)
관심분야 추천시스템, 모바일비즈니스, 고객관계관리, 데이터마이닝, 소셜네트워크 등

김은홍 (E-mail : ehkim@kookmin.ac.kr)
1975년~1979년 서울대학교 경영학과 (경영학학사)
1979년~1981년 KAIST 산업공학과 (석사)
1981년~1986년 KAIST 경영과학 (박사)
1986년~현재 국민대 경영정보학부 교수
관심분야 정보시스템 전략, 정보시스템 조직 및 인력관리, 정보산업 정책

김명균 (E-mail : mkkim@kookmin.ac.kr)
1979년~1983년 국민대학교 경영대학 (학사)
1983년~1987년 KDI 연구원
1987년~1992년 The University of Michigan at Ann Arbor 박사 (재무전공)
1992년~1993년 삼성경제연구소 선임연구원 및 금융팀장
1993년~현재 국민대학교 경영학부 재무전공 교수
관심분야 M&A, 경제성 분석, EVA, Valuation model