

네트워크 중심성 척도가 추천 성능에 미치는 영향에 대한 연구*

이동원

한성대학교 사회과학부 벤처경영트랙
(dongwonlee@hansung.ac.kr)

개인화 추천에서 많이 사용되는 협업 필터링은 고객들의 구매이력을 기반으로 유사고객을 찾아 상품을 추천할 수 있는 매우 유용한 기법으로 인식되고 있다. 그러나, 전통적인 협업 필터링 기법은 사용자 간에 직접적인 연결과 공통적인 특징을 기반으로 유사도를 계산하는 방식으로 인해 신규 고객 혹은 상품에 대해 유사도를 계산하기 힘들다는 문제가 제기되어 왔다. 이를 극복하기 위하여, 다른 기법을 함께 사용하는 하이브리드 기법이 고안되기도 하였다. 이런 노력의 하나로써, 사회연결망의 구조적 특성을 적용하여 이런 문제를 해결하려는 시도가 있었다. 이는, 직접적으로 유사성을 찾기 힘든 사용자 간에도 둘 사이에 놓인 유사한 사용자 또는 사용자들을 통해 유추해내는 방식으로 상호 간의 유사성을 계산하는 방식을 적용한 것이다. 즉, 구매 데이터를 기반으로 사용자의 네트워크를 생성하고 이 네트워크 내에서 두 사용자를 간접적으로 이어주는 네트워크의 특성을 기반으로 둘 사이의 유사도를 계산하는 것이다. 이렇게 얻은 유사도는 추천대상 고객이 상품의 추천에 대한 수락여부를 결정하는 척도로 활용될 수 있다. 서로 다른 중심성 척도는 추천성과에 미치는 영향이 서로 다를 수 있다는 점에서 중요한 의미를 갖는다 할 수 있다. 이런 유사도의 계산을 위해서 네트워크의 중심성을 활용할 수 있다. 본 연구에서는 여기서 더 나아가 이런 중심성이 추천성과에 미치는 영향이 추천 알고리즘에 따라서도 다를 수 있다는 데에서 주목하여 수행되었다. 또한, 이런 네트워크 분석을 활용한 추천기법은 신규 고객 혹은 상품뿐만 아니라 전체 고객 혹은 상품으로 그 대상을 넓히더라도 추천 성능을 높이는 데 기여할 것을 기대할 수 있을 것이다. 이런 관점에서 본 연구는 네트워크 모형에서 연결선이 생성되는 것을 이진 분류의 문제로 보고, 추천 모형에 적용할 분류 기법으로 의사결정나무, K-최근접이웃법, 로지스틱 회귀분석, 인공신경망, 서포트 벡터 머신을 선택하고, 온라인 쇼핑몰에서 4년2개월 간 수집된 구매 데이터로 실험을 진행하였다. 사회연결망에서 측정된 중심성 척도를 각 분류 기법에 적용하여 생성한 모형을 비교 실험한 결과, 각 모형 별로 중심성 척도의 추천성공률이 서로 다르게 나타남을 확인할 수 있었다.

주제어 : 추천 시스템, 개인화 추천, 사회 연결망 분석, 중심성 척도, 분류 모형

논문접수일 : 2020년 11월 4일 논문수정일 : 2021년 1월 28일 게재확정일 : 2021년 2월 2일

원고유형 : 학술대회용 Fast-Track 교신저자 : 이동원

1. 서론

물리적 제약을 받지 않는 온라인 공간에서 진행되는 상품의 종류는 거의 무한에 가깝게 확장될 수 있게 되었다. 이는 고객의 입장에서는, 자

신의 취향에 맞는 상품의 선택의 폭을 넓혀준 반면, 다양한 대안을 비교해야 하는 노력을 크게 증가시켰다. 모든 대안을 검토하는 것이 현실적으로 불가능한 상황에 놓인 고객의 선택을 돕기 위한 기능으로 온라인 기업은 추천 시스템을 도

* 본 연구는 한성대학교 교내학술연구비 지원과제 임

입하였다.

추천 시스템에는 주로 협업 필터링(Collaborative Filtering)이라는 기법을 활용하게 되는데, 추천 대상이 되는 목표 사용자와 유사한 선호도를 보이는 사용자를 찾아 목표 사용자가 선호한 상품을 목표 고객에게 추천하는 사용자 기반(User-based) 기법과, 추천 대상 상품과 구매 유사도를 보이는 유사 상품을 추천 대상으로 선정하는 아이템 기반(Item-based) 기법이 있다. 이때, 선호도의 표현은 구체적인 평가로부터 직접적으로 확보하거나 구매 이력을 통해 간접적으로 유추할 수 있다. 이 과정에서 목표 고객(또는 상품)과 유사한 고객(또는 상품)을 찾는 것이 추천의 성과에 큰 영향을 미친다. 이 때, 동일한 상품에 대해 선호도를 표현(또는 구매)한 이력이 존재해야 두 사용자 간의 유사도를 계산할 수 있으나, 많은 상품을 취급하는 온라인 쇼핑의 경우 이런 교집합이 형성되지 않는 경우가 많아, 유사도를 계산할 수 없는 사용자 간의 관계가 많이 존재한다. 이런 단점을 보완하기 위해 협업 필터링은 다른 기법과 결합된 하이브리드(Hybrid) 형태를 띄기도 하는데, 상품의 속성을 기반으로 추천 대상을 선정하는 내용 기반 필터링(Content-based Filtering) 기법이 그 대표적인 예라 할 수 있다.

신규 고객이나 상품의 경우 유사 고객이나 상품을 찾기 힘들게 되는 추천 모형의 한계를 극복하기 위한 또 다른 방법으로서 사회연결망을 활용하는 기법도 연구되고 있는데, 이는 고객 간에 간접적인 연결을 통해 연관성을 찾아 유사도를 예측하는 방법이다. 즉, 두 고객 간에는 직접적인 연결이 존재하지 않는 경우 두 고객과 공통으로 연결되는 고객을 통해 둘 사이의 유사도를 간접적으로 계산한다. 이는 고객 또는 상품의 관련

성을 기반으로 구성된 사회연결망이 갖는 구조적 특성으로부터 측정된 척도를 활용하는 방법이다. 그 중에는, 중심성(Centrality)을 이용하는 방법도 연구되고 있는데, 이는 네트워크 내에서 각 노드가 네트워크의 중심에 놓이는 정도를 다양한 기준으로 측정한 결과값으로서, 연결정도 중심성(Degree Centrality), 근접 중심성(Closeness Centrality), 매개 중심성(Betweenness Centrality), 위세 중심성(Eigenvector Centrality)을 포함하여 여러 척도가 개발되었다.

본 연구는 이런 중심성이 추천 시스템에 적용되었을 때 추천 성과에 미치는 영향을 파악하고자 수행되었다. 추천의 성공여부는 사회연결망에서 노드로 표현되는 고객이 상품에 대해 직접적인 링크가 존재하지 않는 상황에서 둘 간에 새로운 링크가 생성될 수 있는 확률의 영향을 받을 것이며, 이는 중심성 척도로부터 일정 수준 예측 가능할 것이기 때문이다. 이는 연결이 되는 경우와 연결이 되지 않는 경우라는 두 값을 갖는 종속변수와 중심성 척도를 독립변수로 갖는 분류 모형으로 표현될 수 있다. 분류에 사용 가능한 모형은 다양하기 때문에 이런 척도를 적용하였을 때 모든 모형에서 보편적으로 높은 성능을 보이는 척도가 있을 수도 있을 것이나, 중심성 척도는 그 특성이 다르므로 인해 이를 적용한 추천의 성과도 서로 다른 알고리즘을 적용하는 모형별로 다른 성과를 보일 수 있다. 이런 경우, 최적의 성능을 얻기 위한 추천 시스템은 선택되는 분류기법에서 최적의 성능을 보이는 척도를 독립변수로 채택한 모형일 것이다. 또한, 이런 추천 기법은 신규 고객 혹은 상품에 국한되지 않고 전체 고객 혹은 상품으로 그 대상을 넓히더라도 추천 성능을 높이는 데 기여할 것을 기대할 수 있을 것이다.

이런 아이디어를 검증하기 위한 분류기법으로 의사결정나무(Decision Tree), K-최근접이웃법(K-Nearest Neighbors), 로지스틱 회귀(Logistic Regression), 인공신경망(Neural Network), 서포트 벡터 머신(Support Vector Machine)을 채택하였다. 개별 중심성 척도가 각 기법에서 보이는 성능을 비교하기 위하여 각 중심성 척도만을 독립 변수로 갖는 모형을 생성하고 이를 온라인 쇼핑몰에서 확보한 데이터로 검증하였다. 데이터는 일정 시점을 기준으로 구분하였으며, 기준 시점 이전 데이터로부터 사회연결망을 생성하고, 이후 시점 데이터로부터 훈련 데이터와 검증 데이터를 생성하여 모형을 훈련하고 성과를 측정하는 과정으로 수행되었다. 수행 결과 각 기법 별로 추천 시스템에 적용하였을 때 더 높은 성과를 낼 수 있는 중심성 척도가 서로 다르다는 것을 확인할 수 있었다.

이후의 구성은 다음과 같다. 2장에서는 이론적 배경을 설명하고, 3장에서는 연구 모형의 설계 과정을 다룬다. 각 중심성 척도를 적용한 추천 모형의 성능을 평가한 결과를 4장에서 상세히 다루고, 5장에서 결론으로 마무리한다.

2. 이론적 배경

2.1. 추천 시스템(Recommender System)

추천 기능은 고객이 원하는 상품을 찾는 탐색 노력을 줄여줄 뿐만 아니라 고객 충성도의 상승으로 인해 매출을 증대시키는 효과를 갖는다(Ansari et al., 2000). 이는 주로 협업 필터링(Collaborative Filtering)이라는 기법을 활용하게 되는데, 추천 대상이 되는 목표 사용자와 유사한

선호도를 보이는 사용자를 찾아 목표 사용자가 선호한 상품을 목표 고객에게 추천하는 사용자 기반(User-based) 기법과, 추천 대상 상품과 구매 유사도를 보이는 유사 상품을 추천 대상으로 선정하는 아이템 기반(Item-based) 기법이 있다(Konstan et al., 1997; Ansari et al., 2000). 이 때, 선호도의 표현은 구체적인 평가로부터 직접적으로 확보하거나 구매 이력을 통해 간접적으로 확보할 수 있다. 이 과정에서 목표 고객(또는 상품)과 유사한 고객(또는 상품)을 잘 찾는 것이 추천의 성과에 큰 영향을 미친다. 이 때, 동일한 상품에 대해 선호도를 표현(또는 구매)한 이력이 존재해야 두 사용자 간의 유사도를 계산할 수 있으나, 많은 상품을 취급하는 온라인 쇼핑의 경우 이런 교집합이 형성되지 않는 경우가 많아, 유사도를 계산할 수 없는 사용자 간의 관계가 많이 존재한다. 이를 보완하기 위해 상품의 속성을 기반으로 상품 간의 유사성을 찾는 내용 기반 필터링(Content-based Filtering) 기법이 연구되었고(Konstan et al., 1997; Ansari et al., 2000), 두 가지를 결합한 하이브리드(Hybrid) 형태도 널리 사용되고 있다.

추천과 관련해서는 다양한 연구가 진행 중이다. 추천 성능을 높이기 위한 연구(Balabanovic and Shoham, 1997; Ansari et al., 2000; Adomavicius and Tuzhilin, 2011; Choi et al., 2016), 추천 시스템의 성과를 측정하는 방법에 관한 연구(Bodapati, 2008; Fleder and Hosanagar, 2009), 상거래 이외의 다양한 분야에 적용한 연구(Choi et al., 2015; Kim and Lee, 2013; Kim et al., 2010) 등이 있다. 구체적으로는, 의사결정나무(Cho et al., 2002; Gershman et al., 2010; Kim et al., 2001; Li and Yamada, 2004), 최근접이웃법(Adeniyi et al., 2016; Ahuja et al., 2019; Gemmell et al., 2009; Sarwar et al., 2001; Sarwar et al., 2002;

Subramaniaswamy et al., 2017), 로지스틱 회귀 (Parra et al., 2011; Montanés et al., 2009; Wang et al., 2016), 인공신경망(Katzman et al., 2018; Lee et al., 2002; Paradarami et al., 2017; Christakou et al., 2007; Cheng et al., 2016; Vassiliou et al., 2006; Twardowski, 2016; Devi et al., 2010), 서포트 벡터 머신(Di Noia et al., 2012; Karthikeyan et al., 2017; Min and Han, 2005; Wang et al., 2016), 연관규칙 마이닝(Aher and Lobo, 2012; Duan et al., 2011; Kim and Yum, 2011; Lee, 2017; Lee et al., 2013; Paranjape-Voditel and Deshpande, 2011)와 같이 다양한 기법이 추천 시스템의 성능을 높이기 위해 적용된 바 있다.

2.2. 사회연결망(Social Network)

사회연결망 분석은 사회적 개체가 서로 상호 작용하는 사회적 구조를 파악함으로써, 개체의 개별 속성이 아닌 개체 간의 관계를 이해하려는 방법을 의미한다(Yun and Chae, 2005; Sohn, 2002; Kim, 2003). 노드(Node)로 표현되는 개체와 링크(Link)로 표현되는 이들 간의 관계는 연결망(Network)에서 연결 상태 및 구조를 계량적으로 측정하고 시각적으로 표현된다. 사회연결망을 추천 시스템에 적용하기 위해 다양한 연구(Kim and Chang, 2010; Kim and Kim, 2014; Kim and Kim, 2016; Noh et al., 2017; Kang, 2010; Kim et al., 2010; Lee, 2017; Shin et al., 2012; Part et al., 2009)가 수행되었다.

연결망의 특성은 밀도(Density), 중심화(Centralization), 중심성(Centrality)과 같은 척도로 표현된다. 이 중에서, 밀도는 연결망 내에서 노드들이 얼마나 서로 많이 연결되었는가를 판단하는 척도로서, 연결 가능한 링크의 수에 대해 실제로 연결된 링

크의 수의 비율로 계산된다. 중심화는 특정 노드를 중심으로 연결망이 집중되어 있는 수준을 측정하는 척도이며, 연결 정도 집중도(Degree Centralization), 근접 집중도(Closeness Centralization), 매개 집중도(Betweenness Centralization)가 보편적으로 사용된다. 중심성은 각 노드가 연결망 내에서 중심적인 역할을 수행하는 수준을 측정하는 척도로서, 대표적으로 연결 정도 중심성(Degree Centrality), 근접 중심성 (Closeness Centrality), 매개 중심성 (Betweenness Centrality), 위세 중심성(Eigenvector Centrality) 등이 활용되고 있다. 연결 정도 중심성은 특정 노드에 연결된 다른 노드의 수를 측정하는 척도이다. 근접 중심성은 특정 노드에 연결된 다른 노드들의 거리를 측정하는 척도이다. 매개 중심성은 특정 노드가 다른 노드들 간의 최단 경로에 위치하는 비율을 측정하는 척도이다.

사회연결망을 추천 시스템에 적용하기 위해 다양한 연구가 진행되었다. Park et al.(2009)은 신규고객 추천문제에 대해 새로운 해법을 제시하고자 사회연결망 분석기법을 적용하였다. 사회연결망의 기존 노드에 새로 연결되는 노드는 무작위가 아닌 선호성을 기반으로 한다(Albert and Barabasi, 2002)는 점을 근거로, 링크가 많은 노드에 새로운 노드가 연결될 가능성이 높으며, 이를 활용하여 구매 가능성을 예측할 수 있다는 점을 추천 시스템에 활용하고자 하였다. 즉, 네트워크의 척도 중 연결 정도 중심성을 활용하여 새로운 링크가 생성될 가능성을 새로운 구매가 일어날 가능성으로 해석하는 방법을 고안하였다. 이후로 네트워크를 추천 모형을 개발하기 위한 다양한 연구(Cho and Bang, 2009; Kang, 2010; Lee et al., 2011; Kang and Kwahk, 2011; Cho and Bang, 2011; Ahn et al., 2012; Kim and Kim, 2014; Choi et al., 2016; Noh and Ahn, 2017; Lee, 2017)

가 이루어졌다. Cho and Bang(2009)은 이런 네트워크의 중심성 중 연결 정도 중심성뿐만 아니라 근접 중심성, 매개 중심성, 위세 중심성을 적용함으로써 이들 간의 성능 차이가 있음을 보였다. 즉, 동일한 추천 모형을 통해 생성된 추천 항목이 각 중심성을 적용함에 따라 서로 다른 확률로 구매됨을 보인 것이다. 본 연구에서는 이런 추천 성능의 차이가 중심성의 종류에 따라 달라질 수 있을 뿐만 아니라, 이를 활용하는 모형에 따라서도 달라질 수 있다는 점을 확인하고자 수행되었다. 예를 들어, 연결 정도 중심성이 하나의 모형에서 좋은 성능을 보인다고 하더라도, 다른 모형에서는 낮은 성능을 보일 수 있고, 대신 근접 중심성이 더 높은 성과를 보일 수 있다는 것이다. 또한, 기존 연구와 달리 신규고객이나 신규상품에 한정하지 않고 기존의 고객과 상품까지 확장함으로써 중심성 척도가 영향을 미칠 수 있는 범위를 넓히고자 한다. 본 연구는 이와 같이 중심성의 종류와 모형의 종류에 따라 어떤 조합이 더 높은 성능을 낼 수 있을지를 확인하기 위한 목적으로 수행된다.

2.3. 분류 모형(Classification)

기존 연구를 살펴보면 네트워크 모형을 활용한 추천 모형에서는 링크의 연결여부를 구매여부로 확인하는 방법으로 추천 성능을 평가하였다. 즉, 링크가 연결될 가능성이 높은 경우 추천이 성공할 확률도 높다고 판단한 것이다. 이는 링크의 연결여부를 분류하는 이진분류의 문제로 해석할 수 있다. 따라서 본 연구는 이런 분류 모형 중 이진분류에 활용되는 빈도가 높은 모형을 선별하여 이를 중심으로 모형과 네트워크 중심성 척도 간 성능비교를 실시한다. 분류 모형으로

는 의사결정나무, K-최근접이웃법, 로지스틱 회귀, 인공신경망, 서포트 벡터 머신을 선정하였다.

2.3.1. 의사결정나무(Decision Tree)

의사결정나무는 분류 및 회귀 분석에 사용되는 비모수(Non-Parametric) 감독 학습(Supervised Learning) 방법으로서, 통계(Statistics), 데이터 마이닝(Data Mining), 기계학습(Machine Learning) 분야에서 널리 사용된다. 데이터가 갖는 여러 입력 변수로부터 단순한 의사결정 규칙을 학습해 목표 변수의 값을 예측하는 모델을 만들기 위한 목적으로 사용된다. 의사결정나무 중 목표 변수가 유한한 수의 값을 갖는 분류 트리와 연속적인 무한한 값을 갖는 회귀 트리로 나뉜다. 의사결정나무는 시각화할 수 있고, 이해하기 쉬우며, 해석하기 쉽다는 장점을 갖는다. 또한, 보통 범주형과 연속형 변수 중 하나에 특화되는 다른 기법들과는 달리 두 가지 유형의 변수를 모두 다룰 수 있다. 또한, 목표 변수의 값을 결정하는 과정이 관찰 가능한 화이트 박스 모델로서, 결과를 해석하는 것이 상대적으로 쉽다. 반면, 지나치게 복잡한 구조의 트리가 생성되어 데이터를 일반화시키지 못하는 과적합(Overfitting) 문제가 발생할 수 있는데, 이는 가지치기(Pruning), 리프 노드(Leaf Node)의 최소 샘플 수 설정, 트리의 최대 깊이 설정과 같은 방법으로 극복 가능하다.

2.3.2. K-최근접이웃법(K-Nearest Neighbors)

K-최근접이웃법은 분류 대상 데이터의 이웃에 기반하여 분류(Classification)와 회귀분석(Regression)에 사용되는 비모수(Non-Parametric) 기법이다. 모델을 구성하는 대신, 훈련 데이터의 인스턴스만을 저장하고, 한 점에 대해 출력값을 정할 때

가장 가까운 양의 정수 K 개의 이웃의 속성값을 사용하게 된다. 분류에 사용하게 될 때에는 가장 가까운 K 개의 이웃 중 다수를 차지하는 클래스로 할당하게 되며, 회귀분석에 적용될 때에는 가장 가까운 K 개의 이웃의 속성에 대해 평균을 구하여 이를 미지의 속성값으로 할당하게 된다.

2.3.3. 로지스틱 회귀(Logistic Regression)

로지스틱 회귀는 기본적인 형태에서 로지스틱 함수를 사용하여 이항 종속 변수를 모형화하는 통계적 모델이다. 회귀 분석에서 로지스틱 회귀 분석은 로지스틱 모형의 모수를 추정하는 것이다. 흔히 로지스틱 회귀분석은 종속변수가 이항형인 문제를 다루는 모형을 언급하는데, 종속변수가 두 개 이상의 범주를 갖는 경우도 가능하며 이를 다항 로지스틱 회귀모형(Multinomial Logistic Regression)이라고 한다. 이항 로지스틱 모형은 성공/실패와 같이 두 개의 값을 갖는 종속 변수를 가지고 있으며, 여기서 두 값은 "0"과 "1"로 표시된다. 독립 변수는 이항 변수나 연속 변수 모두 가능하다.

2.3.4. 인공신경망(Neural Network)

인공신경망은 인간의 뇌를 구성하는 기본단위인 뉴런의 신경망을 모형화한 형태를 지닌다. 뇌의 뉴런이 시냅스 구조로 연결된 것처럼 인공 뉴런이라 불리는 노드들은 다른 노드들에 신호를 전송할 수 있고, 신호를 수신한 뉴런은 이를 처리한 후 또 다시 연결된 뉴런에 신호를 전달하는 과정을 수행한다. 여기서, 처리 과정에서는 실수 값을 갖는 신호를 받아들여, 활성화 함수라고 불리는 비선형 함수의 계산에 의해 출력값을 계산하게 된다. 노드와 이들 간의 연결은 가중치를

갖고 있으며 학습이 진행됨에 따라 이 값은 증가되거나 감소되는 과정이 이루어진다. 인공신경망은 여러 개의 노드로 구성된 계층형 구조를 갖는데, 입력층(Input Layer), 은닉층(Hidden Layer), 출력층(Output Layer)로 구성된다. 은닉층이 2개 이상인 경우 심층 신경망(Deep Neural Network)이라 하며 이를 학습하여 모형을 만드는 것을 흔히 딥러닝(Deep Learning)이라 부른다.

2.3.5. 서포트 벡터 머신(Support Vector Machine)

서포트 벡터 머신(Support Vector Machine, SVM)은 기계 학습의 분야 중 하나로서, 주로 분류와 회귀 분석을 위해 사용되는 지도 학습(Supervised Learning) 모형이다. 서포트 벡터 머신은 학습용 데이터를 비선형 매핑을 통해 고차원으로 변환하고, 여기서 데이터들을 서로 다른 범주로 분리하는 데 최적화된 초평면(Hyperplane)을 찾는다. 이런 초평면을 찾기 위해 서포트 벡터 머신은 서포트 벡터(Support Vector)와 마진(Margin)을 사용한다. 서포트 벡터 머신은 선형적으로 분리되지 않는 경우에 선형 서포트 벡터 머신을 확장하여 비선형 서포트 벡터 머신을 만들 수 있으며, 이를 위해서 주어진 데이터를 고차원 특징 공간으로 사상하는 작업을 해야 하며, 이 과정에서의 계산에 따르는 비용을 줄이기 위해 커널 함수(Kernel Function)를 적용하기도 한다.

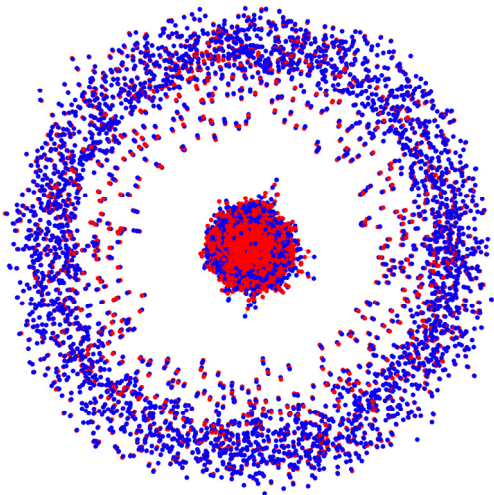
3. 연구 모형

3.1. 데이터 수집

본 연구에서는 고객을 대상으로 다양한 카테

고리의 상품을 판매하는 온라인 상거래 기업으로부터 수집한 실제 주문 거래 데이터를 사용한다. 데이터의 수집 기간은 2012년 4월부터 2016년 4월까지이며, 전체 카테고리를 대상으로 분석을 실시한다. 이들 중, 2016년 1월 이전의 데이터와 이후의 데이터로 구분하고, 전자는 사회연결망을 구성하기 위한 용도로, 후자는 예측모형을 학습하고 검증하기 위한 용도로 사용한다.

분석에 사용된 상품의 수는 19,194개이며, 이를 구매한 고객의 수는 모두 97,223명이다. 사회연결망을 구성하는 노드는 고객과 상품으로서, 고객이 상품을 구매한 이력이 있는 경우 둘 사이에 링크를 생성하는 방법을 적용하되, 여러 건의 주문이 있더라도 가중치는 1만 부여하는 방식으로 <Figure 1>과 같은 고객-상품 주문 연관성 네트워크를 생성한다. <Figure 1>에서 고객과 상품은 각각 빨간색과 파란색으로 구분된다. 사회연결망의 분석은 Python의 NetworkX 패키지를 활용한다.



<Figure 1> Customer-Product Network of Order Records

3.2. 모형 설계 및 분석

목표 고객이 목표 상품을 주문할 것인지를 예측하는 모형을 작성하기 위해 사회연결망 분석 기법은 다음과 같이, 예측에 사용되는 데이터로부터 고객과 상품의 주문여부에 해당하는 데이터를 생성하고, 이들을 포함하는 사회연결망을 구성한 후, 예측모형을 학습시키는 과정으로 진행된다.

3.2.1. 추천대상 선정

예측 목적을 위한 데이터로부터 고객, 상품 별 구매 여부에 해당하는 레코드를 생성한다. 구매 여부를 결정하는 고객과 상품은 사회연결망으로부터 중심성을 계산할 수 있어야 하므로, 사회연결망 생성을 위한 데이터에 포함된 고객과 상품으로 한정한다. 또한, 목표 상품에 대해 목표 고객의 구매여부를 예측하는 상황이므로 사회연결망 데이터에서 해당 고객이 주문건이 발생하지 않은 상품으로 추천대상을 한정한다.

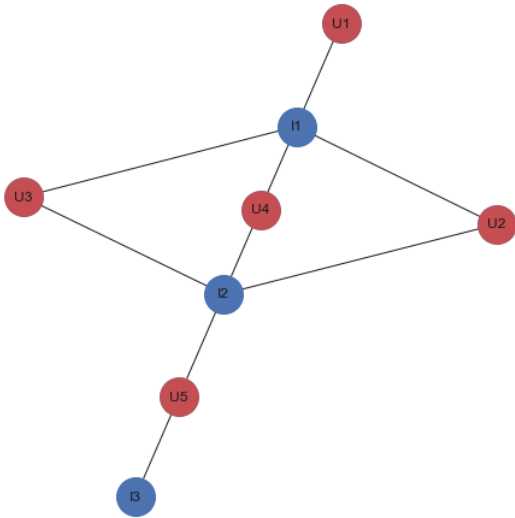
3.2.2. 사회연결망 구성

목표 고객과 목표 상품 간의 최단경로를 찾아 이를 구성하는 노드와 링크로부터 네트워크를 생성한다. 예를 들어, 고객(Customer)와 상품(Product)에 대해 <Table 1>과 같은 주문 거래 내역이 있다고 가정한다.

<Table 1> Order Records

Customer	Product
C1	I1
C2	I1, I2
C3	I1, I2
C4	I1, I2
C5	I2, I3

이 때, U1에게 I3를 추천하고자 한다면 U1에서 I3에 도달하는 경로에 놓인 모든 노드와 링크를 포함하는 서브 네트워크를 그리면 <Figure 2>와 같다.



<Figure 2> 2-Mode Customer-Product Network

이 때, 고객 U1이 구매한 경험이 없는 상품 I3를 추천에 의해 주문할 것인지를 다음과 같은 방법으로 예측한다. U1은 I3에 대한 구매이력이 없으므로 직접 연결된 링크가 없는 상황에서, U1과 I3를 간접적으로 연결하는 네트워크에 의해 둘 사이의 직접적인 연결의 가능성이 결정된다고 가정한다. 즉, U1과 I3가 함께 속한 네트워크의 구조적 특성을 나타내는 독립 변수로부터 둘 간의 링크의 향후 생성 여부라는 종속변수를 예측하는 문제로 볼 수 있다. 이런 네트워크의 특성을 알려주는 지표로서 본 연구에서는 중심성(Centrality) 척도를 적용하도록 한다.

3.2.3. 중심성 계산

네트워크로부터 얻을 수 있는 중심성은 다양하나, 그 중 보편적으로 사용되는 연결 중심성(Degree Centrality), 근접 중심성(Closeness Centrality), 매개 중심성(Betweenness Centrality), 위세 중심성(Eigenvector Centrality)을 계산한다. 각 중심성은 네트워크에 속한 각 노드에 대해 계산이 가능하며, 본 연구에서는 네트워크의 특성을 대표하는 값으로서 네트워크에 속한 모든 노드의 중심성에 대해 평균값을 사용하도록 한다.

이와 같은 방식으로 <Figure 2>에 표현된 사회 연결망으로부터 중심성 척도를 계산하면, 연결 정도 중심성은 0.3214, 근접 중심성은 0.4948, 매개 중심성은 0.1845, 위세 중심성은 0.3249을 얻을 수 있다.

3.2.4. 모형 학습

데이터마이닝의 분류기법 중 활용도가 높은 모형을 추천모형으로 학습과 검증을 실시한다. 본 연구에서는 의사결정나무, K-최근접이웃법, 로지스틱 회귀, 인공신경망, 서포트 벡터 머신을 선택하였다.

앞서 계산된 중심성 척도들을 적용하여 모형을 학습시킨다. 예측용 데이터는 학습용과 검증용을 8:2의 비율로 임의 추출하는 방법으로 구성하였다. 사회연결망에서 생성된 중심성 척도들을 적용하여 학습용 데이터를 통해 모형을 학습시킨 후 검증용 데이터로 검증하여 모형의 분류 정확도를 계산한다. 학습용 데이터에서는 고객과 상품에 대한 짝을 생성하되 실제로 구매한 경우는 1, 구매하지 않은 경우는 0으로 구매여부를 설정한다. 이렇게 설정된 값은 추천 시 고객이 수락하는지 여부를 학습시키기 위한 종속변수로

<Table 2> Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
RecommAccept	40038	0.5000	0.5000	0.0000	1.0000
Degree Cent.	40038	0.1416	0.1402	0.0000	0.5000
Closeness Cent.	40038	0.3244	0.1554	0.0000	0.6410
Betweenness Cent.	40038	0.1147	0.1179	0.0000	0.3333
Eigenvector Cent.	40038	0.1792	0.1459	0.0000	0.5000

<Table 3> Correlation Coefficients

	RecommAccept	DegreeCent	ClosenessCent	BetweennessCent	EigenvectorCent
RecommAccept	1.0000				
Degree Cent.	0.0753	1.0000			
Closeness Cent.	0.1420	0.7907	1.0000		
Betweenness Cent.	0.0276	0.9289	0.6843	1.0000	
Eigenvector Cent.	0.0094	0.0004	-0.0043	-0.0014	1.0000

활용된다. 앞서 계산된 중심성을 적용한 분류모형은 아래와 같이 표현할 수 있다.

$$RecommAccept = \beta_0 + \beta_1 Cent + \varepsilon \quad (1)$$

여기서, *RecommAccept*는 추천수용여부(Recommendation Acceptance), *Cent*는 중심성으로서, 연결 정도 중심성(Degree Centrality), 근접 중심성(Closeness Centrality), 매개 중심성(Betweenness Centrality)를 의미한다. 본 연구에서 사용된 변수는 Python의 NetworkX 패키지로 부터 계산되었으며, <Table #>과 <Table #>는 각각 이들 변수의 기술통계량과 변수 간 상관계수를 보여준다. <Table #>에서 볼 수 있듯이 각 독립변수들 간의 상관관계가 매우 높아 적절한 중심성 척도를 선별하는 것이 중요함을 알 수 있다.

4. 성능 평가

3장에서 분석한 모형의 성능은 다음과 같이 평가한다. 모형의 평가는 혼동행렬(Confusion Matrix)을 구하고 이로부터 F1-Score를 구하여 비교함으로써 이뤄진다. 즉, 모형 별로 각 중심성에 대해 추천 시 구매여부를 예측하고 이것이 실제 데이터와 일치하는 경우 ‘성공’, 그렇지 않은 경우 ‘실패’의 두 가지로 클래스를 나누어 이에 대한 결과를 혼동행렬로 만들고 이로부터 성과지표인 F1-Score를 계산한다. 추천에 사용되는 각 분류모형의 검증 데이터에 대한 분류결과는 <Table 4>, <Table 5>, <Table 6>, <Table 7>, <Table 8>에 혼동행렬(Confusion Matrix)의 형태로 표시되어 있다.

〈Table 4〉 Logistic Regression's Confusion Matrix

		Predicted	
		0	1
Actual	0	2,306	1,722
	1	1,971	2,009

		Predicted	
		0	1
Actual	0	1,878	2,150
	1	1,285	2,695

		Predicted	
		0	1
Actual	0	1,984	2,044
	1	1,520	2,460

		Predicted	
		0	1
Actual	0	1,532	2,496
	1	1,563	2,417

〈Table 5〉 Support Vector Machine's Confusion Matrix

		Predicted	
		0	1
Actual	0	2,711	1,317
	1	2,474	1,506

		Predicted	
		0	1
Actual	0	749	3,279
	1	397	3,583

		Predicted	
		0	1
Actual	0	2,947	1,081
	1	2,408	1,572

		Predicted	
		0	1
Actual	0	1,594	2,434
	1	1,610	2,370

〈Table 6〉 K-Nearest Neighbors' Confusion Matrix

		Predicted	
		0	1
Actual	0	2,639	1,389
	1	2,150	1,830

		Predicted	
		0	1
Actual	0	1,931	2,097
	1	1,640	2,340

		Predicted	
		0	1
Actual	0	2,677	1,351
	1	2,058	1,922

		Predicted	
		0	1
Actual	0	2,327	1,701
	1	2,222	1,758

<Table 7> Neural Network' s Confusion Matrix

(a) Degree Centrality				(b) Closeness Centrality			
		Predicted				Predicted	
		0	1			0	1
Actual	0	3,092	936	Actual	0	2,730	1,298
	1	2,449	1,531		1	2,030	1,950

(c) Betweenness Centrality				(d) Eigenvector Centrality			
		Predicted				Predicted	
		0	1			0	1
Actual	0	2,132	1,896	Actual	0	684	3,344
	1	1,505	2,475		1	672	3,308

<Table 8> Decision Tree' s Confusion Matrix

(a) Degree Centrality				(b) Closeness Centrality			
		Predicted				Predicted	
		0	1			0	1
Actual	0	1,811	2,217	Actual	0	2,618	1,410
	1	1,290	2,690		1	1,876	2,104

(c) Betweenness Centrality				(d) Eigenvector Centrality			
		Predicted				Predicted	
		0	1			0	1
Actual	0	1,951	2,077	Actual	0	1,983	2,045
	1	1,346	2,634		1	1,897	2,083

각 중심성 별로 각 분류모형의 성능을 <Table 9>, <Table 10>, <Table 11>, <Table 12>, <Table 13>에 표시하였고, 이 중에서 성능의 비교를 위

해 보편적으로 쓰이는 F1-Score를 <Table 14>에 별도로 표시하였다.

<Table 9> Logistic Regression Model' s Classification Report

		0	1	accuracy	macro avg	weighted avg
Degree Centrality	precision	0.5392	0.5385	0.5388	0.5388	0.5388
	recall	0.5725	0.5048	0.5388	0.5386	0.5388
	f1-score	0.5553	0.5211	0.5388	0.5382	0.5383
	support	4028.0000	3980.0000	0.5388	8008.0000	8008.0000
Closeness Centrality	precision	0.5937	0.5562	0.5711	0.5750	0.5751
	recall	0.4662	0.6771	0.5711	0.5717	0.5711
	f1-score	0.5223	0.6108	0.5711	0.5665	0.5663
	support	4028.0000	3980.0000	0.5711	8008.0000	8008.0000

		0	1	accuracy	macro avg	weighted avg
Betweenness Centrality	precision	0.5662	0.5462	0.5549	0.5562	0.5563
	recall	0.4926	0.6181	0.5549	0.5553	0.5549
	f1-score	0.5268	0.5799	0.5549	0.5534	0.5532
	support	4028.0000	3980.0000	0.5549	8008.0000	8008.0000
Eigenvector Centrality	precision	0.4950	0.4920	0.4931	0.4935	0.4935
	recall	0.3803	0.6073	0.4931	0.4938	0.4931
	f1-score	0.4302	0.5436	0.4931	0.4869	0.4865
	support	4028.0000	3980.0000	0.4931	8008.0000	8008.0000

〈Table 10〉 Support Vector Machine Model' s Classification Report

		0	1	accuracy	macro avg	weighted avg
Degree Centrality	precision	0.5229	0.5335	0.5266	0.5282	0.5281
	recall	0.6730	0.3784	0.5266	0.5257	0.5266
	f1-score	0.5885	0.4427	0.5266	0.5156	0.5161
	support	4028.0000	3980.0000	0.5266	8008.0000	8008.0000
Closeness Centrality	precision	0.6536	0.5222	0.5410	0.5879	0.5883
	recall	0.1859	0.9003	0.5410	0.5431	0.5410
	f1-score	0.2895	0.6609	0.5410	0.4752	0.4741
	support	4028.0000	3980.0000	0.5410	8008.0000	8008.0000
Betweenness Centrality	precision	0.5503	0.5925	0.5643	0.5714	0.5713
	recall	0.7316	0.3950	0.5643	0.5633	0.5643
	f1-score	0.6282	0.4740	0.5643	0.5511	0.5515
	support	4028.0000	3980.0000	0.5643	8008.0000	8008.0000
Eigenvector Centrality	precision	0.4975	0.4933	0.4950	0.4954	0.4954
	recall	0.3957	0.5955	0.4950	0.4956	0.4950
	f1-score	0.4408	0.5396	0.4950	0.4902	0.4899
	support	4028.0000	3980.0000	0.4950	8008.0000	8008.0000

〈Table 11〉 K-Nearest Neighbors Model' s Classification Report

		0	1	accuracy	macro avg	weighted avg
Degree Centrality	precision	0.5511	0.5685	0.5581	0.5598	0.5597
	recall	0.6552	0.4598	0.5581	0.5575	0.5581
	f1-score	0.5986	0.5084	0.5581	0.5535	0.5538
	support	4028.0000	3980.0000	0.5581	8008.0000	8008.0000
Closeness Centrality	precision	0.5407	0.5274	0.5333	0.5341	0.5341
	recall	0.4794	0.5879	0.5333	0.5337	0.5333
	f1-score	0.5082	0.5560	0.5333	0.5321	0.5320
	support	4028.0000	3980.0000	0.5333	8008.0000	8008.0000

		0	1	accuracy	macro avg	weighted avg
Betweenness Centrality	precision	0.5654	0.5872	0.5743	0.5763	0.5762
	recall	0.6646	0.4829	0.5743	0.5738	0.5743
	f1-score	0.6110	0.5300	0.5743	0.5705	0.5707
	support	4028.0000	3980.0000	0.5743	8008.0000	8008.0000
Eigenvector Centrality	precision	0.5115	0.5082	0.5101	0.5099	0.5099
	recall	0.5777	0.4417	0.5101	0.5097	0.5101
	f1-score	0.5426	0.4726	0.5101	0.5076	0.5078
	support	4028.0000	3980.0000	0.5101	8008.0000	8008.0000

〈Table 12〉 Neural Network Model's Classification Report

		0	1	accuracy	macro avg	weighted avg
Degree Centrality	precision	0.5580	0.6206	0.5773	0.5893	0.5891
	Recall	0.7676	0.3847	0.5773	0.5762	0.5773
	f1-score	0.6463	0.4750	0.5773	0.5606	0.5611
	support	4028.0000	3980.0000	0.5773	8008.0000	8008.0000
Closeness Centrality	precision	0.5735	0.6004	0.5844	0.5869	0.5869
	Recall	0.6778	0.4900	0.5844	0.5839	0.5844
	f1-score	0.6213	0.5396	0.5844	0.5804	0.5807
	support	4028.0000	3980.0000	0.5844	8008.0000	8008.0000
Betweenness Centrality	precision	0.5862	0.5662	0.5753	0.5762	0.5763
	Recall	0.5293	0.6219	0.5753	0.5756	0.5753
	f1-score	0.5563	0.5927	0.5753	0.5745	0.5744
	support	4028.0000	3980.0000	0.5753	8008.0000	8008.0000
Eigenvector Centrality	precision	0.5044	0.4973	0.4985	0.5009	0.5009
	Recall	0.1698	0.8312	0.4985	0.5005	0.4985
	f1-score	0.2541	0.6223	0.4985	0.4382	0.4371
	support	4028.0000	3980.0000	0.4985	8008.0000	8008.0000

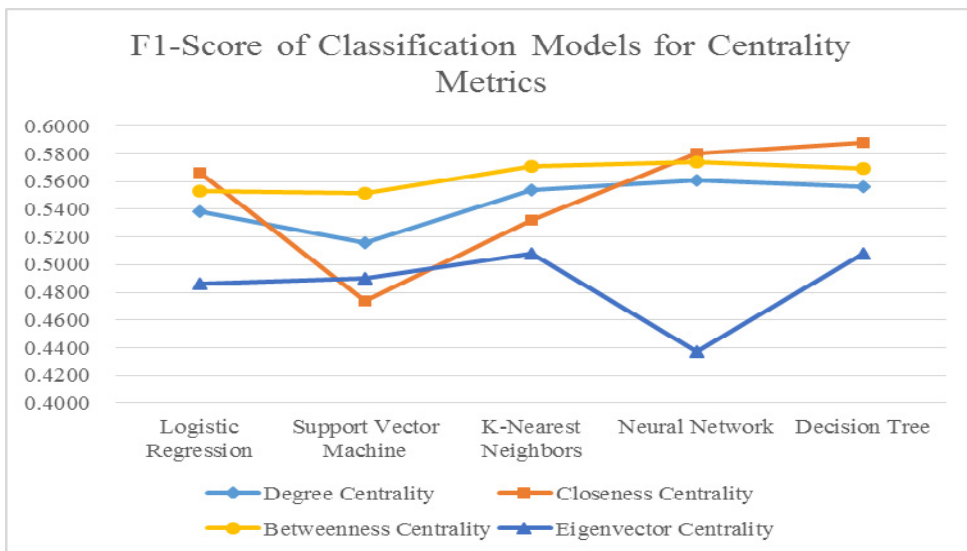
〈Table 13〉 Decision Tree Model's Classification Report

		0	1	accuracy	macro avg	weighted avg
Degree Centrality	precision	0.5840	0.5482	0.5621	0.5661	0.5662
	recall	0.4496	0.6759	0.5621	0.5627	0.5621
	f1-score	0.5081	0.6054	0.5621	0.5567	0.5564
	support	4028.0000	3980.0000	0.5621	8008.0000	8008.0000
Closeness Centrality	precision	0.5826	0.5987	0.5897	0.5907	0.5906
	recall	0.6500	0.5286	0.5897	0.5893	0.5897
	f1-score	0.6144	0.5615	0.5897	0.5880	0.5881
	support	4028.0000	3980.0000	0.5897	8008.0000	8008.0000

		0	1	accuracy	macro avg	weighted avg
Betweenness Centrality	precision	0.5918	0.5591	0.5726	0.5754	0.5755
	recall	0.4844	0.6618	0.5726	0.5731	0.5726
	f1-score	0.5327	0.6061	0.5726	0.5694	0.5692
	support	4028.0000	3980.0000	0.5726	8008.0000	8008.0000
Eigenvector Centrality	precision	0.5111	0.5046	0.5077	0.5078	0.5079
	recall	0.4923	0.5234	0.5077	0.5078	0.5077
	f1-score	0.5015	0.5138	0.5077	0.5077	0.5076
	support	4028.0000	3980.0000	0.5077	8008.0000	8008.0000

<Table 14> F1-Score of Classification Models for Centrality Metrics

	Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
Logistic Regression	0.5383	0.5663	0.5532	0.4865
Support Vector Machine	0.5161	0.4741	0.5515	0.4899
K-nearest Neighbors	0.5538	0.5320	0.5707	0.5078
Neural Network	0.5611	0.5807	0.5744	0.4371
Decision Tree	0.5564	0.5881	0.5692	0.5076



<Figure 3> F1-Score of Classification Models for Centrality Metrics

<Table 14>의 값을 각 모형과 중심성 별로 비교하기 위해 <Figure 3>에 도식화하였다. 이를 통해 확인할 수 있었던 사항은 다음과 같다. 위

세 중심성(Eigenvector Centrality)의 경우 서포트 벡터 머신을 제외한 모든 모형에서 가장 저조한 성능을 나타냄을 확인할 수 있다. 근접 중심성과

연결 중심성의 경우 모든 모형에 걸쳐 비슷한 성능을 보이고 있는데, 연결 중심성(Degree Centrality)은 여러 모형에 걸쳐 전반적으로 중간 정도의 순위를 차지하는 반면, 매개 중심성(Betweenness Centrality)의 경우에는 비교적 이보다 높은 순위로서 항상 연결 중심성(Degree Centrality)보다는 높은 순위를 보이는 특징을 보인다는 것을 확인할 수 있다. 마지막으로, 근접 중심성(Closeness Centrality)은 모형에 따른 성능의 차이가 확연하게 드러나는 특징을 보이고 있는데, 로지스틱 회귀, 인공지능경망, 의사결정나무에서 1위를 차지할 뿐 아니라 수치상으로도 높은 성능을 확인할 수 있다. 그러나, 서포트 벡터 머신과 K-최근접이웃법에서 매우 낮은 순위를 기록할 뿐만 아니라, 수치상으로도 낮은 수준의 성과를 보인다.

성능 평가의 결과를 통해 알 수 있듯이, 분류 모형을 통해 사회연결망에 속한 두 노드 간의 연결가능성을 예측하는 방법에서 두 노드를 연결시키는 서브네트워크 상에서의 중심성 척도는 유용하게 사용될 수 있고, 분류모형의 종류에 따라 다른 성능을 가져올 수 있다는 점을 확인할 수 있었다. 따라서, 적절한 척도를 선택함으로써 이를 추천시스템에 적용하였을 때에 더 높은 성능을 보이는 결과를 얻을 것으로 기대할 수 있을 것이다. 보편적으로 어떤 모형에서든지 높은 수준의 성능을 기대하는 경우 매개 중심성을 적용하는 방법이 적합할 수 있고, 특정 모형에 한정하여 더 높은 성능을 원하는 경우에는 근접 중심성을 선택적으로 적용하는 것이 바람직한 방법으로 고려될 수 있다고 생각된다.

5. 결론

추천은 목표 고객이 아직 선호도를 표현하지 않은 상품 중 선호할 만한 상품을 선별하는 과정이라 할 수 있고, 선호도를 예측하기 위해 목표 고객과 동일 상품을 구매한 유사 고객을 찾는 것이 필요하다. 그러나, 유사 고객 중에 목표 상품을 구매한 이력이 존재하지 않는 경우 선호도는 다른 고객 또는 상품을 통한 간접적인 방식으로 계산될 수밖에 없다. 이 때, 사회연결망은 이들 고객 또는 상품 간의 관계를 네트워크 형태로 표현해줄 수 있고, 이를 분석한 지표는 그 간접적인 연결의 특성을 수치화하여 추천성가를 예측하는 데에 적용할 수 있다는 점에서 유용하게 활용될 수 있다. 특히, 이들 척도 중 중심성은 목표 고객과 목표 상품을 연결하는 여러 고객과 상품으로 구성된 네트워크의 특성을 잘 표현해줄 수 있는 척도로 활용될 수 있다는 점에서 유용할 것으로 기대할 수 있다.

본 연구에서는 이런 네트워크에 속한 각 노드의 특성이 아닌 이들의 집합체인 네트워크의 특성을 활용하기 위해 대표값으로서 이들의 중심도 평균을 분석에 사용하였다. 중심성 척도는 각각 계산되는 방식이 서로 다르기 때문에 추천모형에 적용 시 서로 다른 성과를 낼 것을 기대할 수 있다. 또한, 추천모형으로 사용되는 분류기법의 알고리즘의 차이로 인해 이들 척도의 성과가 다를 수 있다. 중심성 척도는 추천성능에 다른 성과를 미칠 뿐만 아니라, 이들의 성능 차이는 모형에 따라서도 달라진다는 것을 실험을 통해 확인할 수 있었다.

연결 정도 중심성 척도는 여러 목적으로 활용도가 높은 것에 비해 본 연구에서는 추천 시스템에서의 성과를 높이기 위한 목적으로 큰 두각을

보이지 못했다. 반면에, 근접 중심성과 매개 중심성이 높은 성과를 거둔 양상을 보였다. 이를 해석해보면, 이미 많은 노드들과 연결된 네트워크보다는 각 노드들이 연결된 거리가 짧은 네트워크의 경우에 새로운 링크가 생성될 확률이 높다는 것을 의미하는 것으로 볼 수 있다. 또한, 매개 중심성도 중요한 역할을 하는 것을 볼 수 있는데, 이는 브로커 역할을 하는 노드가 중간에 존재해야 노드들 간의 연결이 원활해지는 결과를 기대할 수 있다는 점에서 높은 성과를 보일 것을 예측해볼 수 있다. 즉, 한 고객이 하나의 상품에 이르는 연결 통로가 되어 주는 노드의 중요성을 보여주는 것이라 생각해볼 수 있다.

전반적으로 연결 중심성보다 근접 중심성과 매개 중심성에서 더 높은 성과를 얻게 된 결과는, 연결 정도 중심성이 가장 큰 역할을 하는 것으로 분석된 기존 연구(조운호와 방정혜, 2009)와 차이를 보이고 있다. 그 원인을 추측해보면, 기존 연구에서는 구매건수가 적은 신상품의 추천이라는 상황을 적용한 것으로 해석해볼 수 있을 것이다. 즉, 중심성 척도로 표현되는 네트워크의 특성뿐만 아니라, 연결되는 대상 노드 자체의 특성인 연결 수(Degree)에 의해서도 영향을 받는다는 것을 의미한다고 생각해볼 수 있다. 이는 어찌 보면 당연한 결과라 할 수 있을 것이며, 추천 모형을 구축할 때에 추천 대상 고객 및 상품의 특성을 반영한 변수를 포함시킴으로써 추천 성과를 향상시킬 수 있을 것이라는 함의를 갖는다고 할 수 있다.

그럼에도 불구하고, 기법에 따라서는 이들 척도가 서로 다른 성과를 보임을 알 수 있었다. 즉, 로지스틱 회귀모형, 인공신경망, 의사결정나무에서 높은 성과를 보이는 근접 중심성이 K-최근접이웃법에서 낮은 성과를 보이며, 더욱이 서포

트 벡터 머신에서는 최하위의 성능을 기록하였다. 모든 모형에서 절대적으로 성능의 우위를 갖는 중심성 척도가 없다는 점은, 각 분류 기법이 가진 특성에 따라 각 중심성 척도의 적합성이 서로 달라진다는 것으로 해석될 수 있을 것이다.

본 연구는 척도를 모형에 따라 다르게 적용함으로써 성능의 차이가 발생할 수 있음을 확인하고, 이를 기반으로 실무에서 추천모형을 개발할 때 도움을 주고자 한 목적으로 수행되었다. 따라서, 분류 모형의 성능의 차이가 어떤 원인에 기인한 것인지는, 변수로 사용된 중심성 척도와 어떤 영향을 미친 것인지를 분류 모형의 알고리즘 차원에서 좀 더 세밀하게 파악해보는 노력이 필요할 것으로 보이므로, 이는 연구의 한계점으로 남기고자 한다.

추천 시스템의 성능을 높이기 위한 목적으로, 중심성 척도를 기존의 다른 변수와 함께 사용하는 방안도 고려해볼 수 있을 것이다. 즉, 중심성 척도 하나만을 추천 시스템에 적용하는 것보다는 다양한 구조적 특성을 나타내는 척도를 함께 적용하여 좀 더 높은 성과를 올릴 수 있는 모형을 찾는 것이 가능할 것이다. 또한, 전통적인 협업 필터링 기법과 결합된 하이브리드 형태로 사용하는 것은 좋은 방안이 될 수 있을 것이라 생각된다. 이와 같이 실제로 작동할 수 있는 추천 모형을 작성하여 성능을 비교하는 실험까지 수행하는 것은 향후 연구과제로 남겨두고자 한다.

본 연구에서는 보편적으로 많이 사용되는 4개의 중심성 척도에 대해서만 분석을 실시하였다. 이는 중심성 척도가 분류 기법 별로 성능에서의 차이를 보임을 확인하기 위한 연구로 수행하기에는 충분한 개수라고 판단되었으나, 사회연결망의 분석에 사용되는 다양한 중심성 척도가 존재하는 만큼, 이를 모두 분석하기에는 한계가 있

있던 측면도 있다. 따라서, 후속 연구에서는 좀 더 다양한 중심성 척도에 대해 성과에 미치는 영향을 분석하는 것이 의미가 있을 것이다. 또한, 중심성 척도도 실제 연결될 두 노드, 즉, 고객과 상품의 또 다른 특성에 따라 다른 성과를 보일 수 있다는 점이 좀 더 구체적으로 확인될 수 있다면 더 의미 있는 결과를 제시할 것으로 기대한다.

참고문헌(References)

- Adeniyi, D. A., Wei, Z., and Yongquan, Y. "Automated web usage data mining and recommendation system using K-Nearest Neighbor (KNN) classification method," *Applied Computing and Informatics*, Vol.12, No.1(2016), 90-108.
- Adomavicius, G., A. Tuzhilin. "Context-Aware Recommender Systems," *Recommender Systems Handbook*, Springer US, (2011), 217~253.
- Aher, S. B., and Lobo, L. M. R. J. "A comparative study of association rule algorithms for course recommender system in e-learning," *International Journal of Computer Applications*, Vol.39, No.1(2012), 48-52.
- Ahuja, R., Solanki, A., and Nayyar, A. "Movie recommender system using K-Means clustering and K-Nearest Neighbor," In *2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence)* (2019, January). 263~268. IEEE.
- Ahn, Sung-Mahn, In Hwan Kim, Byounggu Choi, Yoonho Cho, Eunhong Kim, Myeong-Kyun Kim. (2012). Understanding the Performance of Collaborative Filtering Recommendation through Social Network Analysis. *The Journal of Society for e-Business Studies*, 17(2), 129-147.
- Anand, S.S., A.R. Patrick. "A Data Mining methodology for cross-sales," *Knowledge-Based Systems*, Vol.10, No.7(1998), 449-461.
- Ansari, A., S. Essegai, R. Kohli. "Internet recommender systems," *Journal of Marketing Research*, Vol.37, No.3(2000), 363~375.
- Balabanovic, M., Y. Shoham. "Content-Based, Collaborative, Recommendation," *Communications of the ACM*, Vol.40, No.3(1997), 66~72.
- Bodapati, A.V. "Recommender systems with purchase data," *Journal of Marketing Research*, Vol.45, No.1(2008), 77~93.
- Chen, Y.L., J.M. Chen, C.W. Tung. "A data mining approach for retail knowledge discovery with consideration of the effect of shelf-space adjacency on sales," *Decision Support Systems*, Vol.42, No.3(2006), 1503~1520.
- Cheng, H. T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., Anderson, G., Corrado, G., Chai, W., Ispir M., Anil, R., Haque, Z., Hong, L., Jain, V., Liu, X., and Shar, H. "Wide & deep learning for recommender systems," In *Proceedings of the 1st workshop on deep learning for recommender systems*, (2016, September), 7~10.
- Cho, Yoon Ho, Jae Kyeong Kim, and Soung Hie Kim. "A personalized recommender system based on web usage mining and decision tree induction," *Expert systems with Applications* Vol.23, No.3(2002), 329~342.
- Cho, Yoonho, Jounghae Bang. (2009). Social Network Analysis for New Product Recommendation. *Journal of Intelligence and Information Systems*, 15(4), 183-199.

- Cho, Yoonho, Jounghae Bang. (2011). Applying Centrality Analysis to Solve the Cold-Start and Sparsity Problems in Collaborative Filtering. *Journal of Intelligence and Information Systems*, 17(3), 99-114.
- Choi, S., Hyun, Y., Kim, N. "Improving Performance of Recommendation Systems Using Topic Modeling," *Journal of Intelligence and Information Systems*, Vol.21, No.3(2015), 101~116.
- Choi, Seulbi, Kee-Young Kwahk, Hyunchul Ahn. (2016). Enhancing Predictive Accuracy of Collaborative Filtering Algorithms using the Network Analysis of Trust Relationship among Users. *Journal of Intelligence and Information Systems*, 22(3), 113-127.
- Choi, S., Kwahk, K.-Y., 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), 113~127.
- Christakou, C., Vrettos, S., and Stafylopatis, A. "A hybrid movie recommender system based on neural networks," *International Journal on Artificial Intelligence Tools*, Vol.16, No.5(2007), 771~792.
- Devi, M. K., Samy, R. T., Kumar, S. V., and Venkatesh, P. "Probabilistic neural network approach to alleviate sparsity and cold start problems in collaborative recommender systems," In *2010 IEEE International Conference on Computational Intelligence and Computing Research*, (2010, December). 1~4. IEEE.
- Di Noia, T., Mirizzi, R., Ostuni, V. C., and Romito, D. "Exploiting the web of data in model-based recommender systems," In *Proceedings of the sixth ACM conference on Recommender systems*, (2012, September). 253~256.
- Duan, L., Street, W. N., and Xu, E. "Healthcare information systems: data mining methods in the creation of a clinical recommender system," *Enterprise Information Systems*, Vol.5, No.2(2011), 169~181.
- Fleder, D., K. Hosanagar. "Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity," *Management Science*, Vol.55, No.5(2009), 697~712.
- Gemmell, J., Schimoler, T., Ramezani, M., and Mobasher, B. "Adapting k-nearest neighbor for tag recommendation in folksonomies," In *Proceedings of the 7th International Conference on Intelligent Techniques for Web Personalization & Recommender Systems-Volume 528* (2009, July). 69~80. CEUR-WS. Org.
- Gershman, A., Meisels, A., Lüke, K. H., Rokach, L., Schclar, A., and Sturm, A. "A decision tree based recommender system," *10th International Conference on Innovative Internet Community Systems (I2CS)-Jubilee Edition 2010-* (2010)
- 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, Eun-Young, Kee-Young Kwahk. (2011). Managing Duplicate Memberships of Websites : An Approach of Social Network Analysis. *Journal of Intelligence and Information Systems*, 17(1), 153-169.
- Karthikeyan, P., Selvi, S. T., Neeraja, G., Deepika, R., Vincent, A., and Abinaya, V. "Prevention

- of shilling attack in recommender systems using discrete wavelet transform and support vector machine," In *2016 eighth international conference on Advanced Computing (ICoAC)* (2017, January). 99~104. IEEE.
- Katzman, J. L., Shaham, U., Cloninger, A., Bates, J., Jiang, T., and Kluger, Y. "DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network," *BMC medical research methodology*, Vol.18, No.1(2018), 24.
- Kim, H. K., Choi, I. Y., Ha, K. M., Kim, J. K. "Development of User Based Recommender System using Social Network for u-Healthcare," *Journal of Intelligence and Information Systems*, Vol.16, No.3(2010), 181~199.
- Kim, B. K., S. Lee, S. Bang, J. Kim, and J. H. Lee, "Personalized Recommendation System Using Social Network," *Proceedings of the Conference on Intelligent Information Systems*, Vol.20, No.1(2010), 48~49.
- Kim, J., Lee, S.-W. "The Ontology Based, the Movie Contents Recommendation Scheme, Using Relations of Movie Metadata," *Journal of Intelligence and Information Systems*, Vol.19, No.3(2013), 25~44.
- Kim, J. W., Lee, B. H., Shaw, M. J., Chang, H. L., and Nelson, M. "Application of decision-tree induction techniques to personalized advertisements on internet storefronts," *International Journal of Electronic Commerce*, Vol.5, No.3(2001), 45~62.
- Kim, K.-J., Kim, B.-G. "Product Recommender System for Online Shopping Malls using Data Mining Techniques," *Journal of Intelligence and Information Systems*, Vol.11, No.1(2005), 191~205.
- Kim, M., and K. J. Kim, "Recommender Systems using Structural Hole and Collaborative Filtering," *Journal of Intelligence and Information Systems*, Vol.20, No.4(2014), 107~120.
- Kim, Mingun, Kyoung-jae Kim. (2014). Recommender Systems using Structural Hole and Collaborative Filtering. *Journal of Intelligence and Information Systems*, 20(4), 107-120.
- Kim, M. G., and K. J. Kim, "Recommender Systems using SVD with Social Network Information," *Journal of Intelligence and Information Systems*, Vol.22, No.4(2016), 1~18.
- Kim, S. H., and R. S. Chang, "The Study on the Research Trend of Social Network Analysis and the its Applicability to Information Science," *Journal of the Korean Society for Information Management*, Vol.27, No.4(2010), 71~87.
- Kim, Y. and W.N. Street. "An intelligent system for customer targeting: a data mining approach," *Decision Support Systems*, Vol.37, No.2(2004), 215~228.
- Kim, Y. S., and Yum, B. J. "Recommender system based on click stream data using association rule mining," *Expert Systems with Applications*, Vol.38, No.10(2011), 13320~13327.
- Konstan, J.A., B.N. Miller, D. Maltz, J.L. Herlocker, L.R. Gordon, J. Riedl. "GroupLens: applying collaborative filtering to Usenet news," *Communications of the ACM*, Vol.40, No.3(1997), 77~87.
- Lee, D. "Extension Method of Association Rules Using Social Network Analysis," *Journal of Intelligence and Information Systems*. Vol.23, No.4(2017), 111~126.

- Lee, D., S. Park, and S. Moon. "Utility-based association rule mining: A marketing solution for cross-selling," *Expert Systems with Applications*, Vol.40, No.7(2013), 2715~25.
- Lee, M., Choi, P., and Woo, Y. "A hybrid recommender system combining collaborative filtering with neural network," In *International conference on adaptive hypermedia and adaptive web-based systems*, (2002, May). 531~534. Springer, Berlin, Heidelberg.
- Lee, Seok Kee, Hyeon Jo, Sung Yong Chun. (2011). A Customer Profile Model for Collaborative Recommendation in e-Commerce. JOURNAL OF THE KOREA CONTENTS ASSOCIATION, 11(5), 67-74.
- Li, P., and Yamada, S "A movie recommender system based on inductive learning," In *IEEE Conference on Cybernetics and Intelligent Systems*, Vol.1(2004, December). 318~323. IEEE.
- Min, S. H., and Han, I. "Recommender systems using support vector machines," In *International Conference on Web Engineering*, (2005, July). 387~393. Springer, Berlin, Heidelberg.
- Montanés, E., Quevedo, J. R., Díaz, I., and Ranilla, J. "Collaborative tag recommendation system based on logistic regression," *ECML PKDD Discovery Challenge*, (2009), 173~188.
- Noh, H., S. Choi, and H. Ahn, "Social Network-based Hybrid Collaborative Filtering using Genetic Algorithms," *Journal of Intelligence and Information Systems*, Vol.23, No.2(2017), 19~38.
- Noh, Heeryong, Hyunchul Ahn. (2017). A Study on the Recommendation Algorithm based on Trust/Distrust Relationship Network Analysis. *Journal of Information Technology Applications & Management*, 24(1), 169-185.
- Paradarami, T. K., Bastian, N. D., and Wightman, J. L. "A hybrid recommender system using artificial neural networks," *Expert Systems with Applications*, 83(2017), 300~313.
- Paranjape-Voditel, P., and Deshpande, U. "A stock market portfolio recommender system based on association rule mining," *Applied Soft Computing*, Vol.13, No.2(2013), 1055~1063.
- Parra, D., Karatzoglou, A., Amatriain, X., and Yavuz, I. "Implicit feedback recommendation via implicit-to-explicit ordinal logistic regression mapping," *Proceedings of the CARS-2011*, 5. (2011).
- 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.
- Shin, C. H., J. W. Lee, H. N. Yang, and I. Y. Choi, "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), 19~42.
- Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. "Item-based collaborative filtering recommendation algorithms," In *Proceedings of the 10th international conference on World Wide Web*, (2001, April). 285~295.
- Sarwar, B. M., Karypis, G., Konstan, J., and Riedl, J. "Recommender systems for large-scale e-commerce: Scalable neighborhood formation using clustering," In *Proceedings of the fifth international conference on computer and*

- information technology*, Vol.1 (2002, December). 291~324.
- Sohn D., *Social Network Analysis*, Kyungmoon Publications, 2002.
- Subramaniaswamy, V., and Logesh, R. "Adaptive KNN based recommender system through mining of user preferences," *Wireless Personal Communications*, Vol.97, No.2(2017), 2229~2247.
- Twardowski, B. "Modelling contextual information in session-aware recommender systems with neural networks," In *Proceedings of the 10th ACM Conference on Recommender Systems*, (2016, September). 273~276.
- Vassiliou, C., Stamoulis, D., Martakos, D., and Athanassopoulos, S. "A recommender system framework combining neural networks & collaborative filtering," In *Proceedings of the 5th WSEAS international conference on Instrumentation, measurement, circuits and systems*, (2006, April). 285~290.
- Wang, X., Luo, F., Qian, Y., and Ranzi, G. "A personalized electronic movie recommendation system based on support vector machine and improved particle swarm optimization," *PloS one*, Vol.11, No.11(2016), e0165868.
- Wang, Y., Feng, D., Li, D., Chen, X., Zhao, Y., and Niu, X. "A mobile recommendation system based on logistic regression and Gradient Boosting Decision Trees," In *2016 International Joint Conference on Neural Networks (IJCNN)*, (2016, July). 1896~1902. IEEE.
- Yun, Y., and S. Chae, *Introduction to Complex Systems*, Samsung Economic Research Institute, 2005.
- Y. Kim, *Social Network Analysis*, Pakyoungsa, 2003.

Abstract

A Study on the Effect of Network Centralities on Recommendation Performance

Dongwon Lee*

Collaborative filtering, which is often used in personalization recommendations, is recognized as a very useful technique to find similar customers and recommend products to them based on their purchase history. However, the traditional collaborative filtering technique has raised the question of having difficulty calculating the similarity for new customers or products due to the method of calculating similarities based on direct connections and common features among customers. For this reason, a hybrid technique was designed to use content-based filtering techniques together. On the one hand, efforts have been made to solve these problems by applying the structural characteristics of social networks. This applies a method of indirectly calculating similarities through their similar customers placed between them. This means creating a customer's network based on purchasing data and calculating the similarity between the two based on the features of the network that indirectly connects the two customers within this network. Such similarity can be used as a measure to predict whether the target customer accepts recommendations. The centrality metrics of networks can be utilized for the calculation of these similarities. Different centrality metrics have important implications in that they may have different effects on recommended performance. In this study, furthermore, the effect of these centrality metrics on the performance of recommendation may vary depending on recommender algorithms. In addition, recommendation techniques using network analysis can be expected to contribute to increasing recommendation performance even if they apply not only to new customers or products but also to entire customers or products. By considering a customer's purchase of an item as a link generated between the customer and the item on the network, the prediction of user acceptance of recommendation is solved as a prediction of whether a new link will be created between them. As the classification models fit the purpose of solving the binary problem of whether the link is engaged or not, decision tree, k-nearest neighbors (KNN), logistic regression, artificial neural network, and support vector machine (SVM) are selected in the research. The data for performance

* Corresponding author: Dongwon Lee
College of Social Sciences, Hansung University
116 Samseongyoro-16gil, Seongbuk-gu, Seoul 02876, Korea
Tel: +82-2-760-4250, Fax: +82-2-760-4482, E-mail: dongwonlee@hansung.ac.kr

evaluation used order data collected from an online shopping mall over four years and two months. Among them, the previous three years and eight months constitute social networks composed of and the experiment was conducted by organizing the data collected into the social network. The next four months' records were used to train and evaluate recommender models. Experiments with the centrality metrics applied to each model show that the recommendation acceptance rates of the centrality metrics are different for each algorithm at a meaningful level. In this work, we analyzed only four commonly used centrality metrics: degree centrality, betweenness centrality, closeness centrality, and eigenvector centrality. Eigenvector centrality records the lowest performance in all models except support vector machines. Closeness centrality and betweenness centrality show similar performance across all models. Degree centrality ranking moderate across overall models while betweenness centrality always ranking higher than degree centrality. Finally, closeness centrality is characterized by distinct differences in performance according to the model. It ranks first in logistic regression, artificial neural network, and decision tree with numerically high performance. However, it only records very low rankings in support vector machine and K-neighborhood with low-performance levels. As the experiment results reveal, in a classification model, network centrality metrics over a subnetwork that connects the two nodes can effectively predict the connectivity between two nodes in a social network. Furthermore, each metric has a different performance depending on the classification model type. This result implies that choosing appropriate metrics for each algorithm can lead to achieving higher recommendation performance. In general, betweenness centrality can guarantee a high level of performance in any model. It would be possible to consider the introduction of proximity centrality to obtain higher performance for certain models.

Key Words : Recommender Systems, Personalized Recommendation, Social Network Analysis, Network Centrality, Classification Model

Received : November 4, 2020 Revised : January 28, 2021 Accepted : February 2, 2021

Corresponding Author : Dongwon Lee

저 자 소개



이 동 원

한양대학교에서 재료공학사, 카이스트 경영대학원에서 경영정보학석사 학위와 경영공학박사 학위를 취득했다. LG CNS에서 시스템 엔지니어로 근무한 바 있고, 한성대학교에 조교수로 재직 중이다. 주요연구분야는 데이터마이닝, 딥러닝, 소비자 행동, 추천 시스템이다.