

## 신상품 추천을 위한 사회연결망분석의 활용

조운호  
국민대학교 경영학부  
(www4u@kookmin.ac.kr)

방정혜  
국민대학교 경영학부  
(bangjh@kookmin.ac.kr)

.....

추천시스템에서 가장 많이 활용되고 있는 협업필터링은 고객들의 과거 구매이력을 기반으로 추천하기 때문에 새로이 출시되는 상품을 추천하는 것이 근본적으로 불가능하다. 이와 같은 협업필터링의 한계점을 극복하기 위하여 많은 연구자들은 추천 대상 고객이 선호하는 상품과 유사한 속성을 가진 상품을 추천하는 내용기반 필터링을 협업필터링과 결합한 하이브리드 추천기법을 제시하였다. 그러나 하이브리드 추천기법은 음악, 영화 등 속성 추출이 용이한 일부 상품의 추천에만 활용될 수 있다는 한계가 있다. 따라서 상품 유형에 관계없이 고객에게 신상품을 효과적으로 추천할 수 있는 새로운 접근방법이 제시될 필요가 있다.

본 연구에서는 사회연결망분석에서 관계 및 구조적 특성을 분석하기 위해 널리 활용되고 있는 중심성 개념을 적용하여 상품간의 구매 관계를 파악한 후 이를 기반으로 신상품을 구매할 가능성이 높은 고객을 찾아 신상품을 추천 방법을 제안한다. 추천 프로세스는 구매 유사도 분석, 상품 네트워크 구성, 중심성 분석, 신상품 추천 등 네 단계 절차로 나뉘어진다. 제시한 추천방법의 성능을 평가하기 위하여 국내 유명 백화점 중의 하나인 H백화점의 구매 데이터를 사용하여 실험하였다.

.....

논문접수일 : 2009년 12월 05일    게재확정일 : 2009년 12월 18일    교신저자 : 방정혜

### 1. 서론 및 배경

추천시스템은 통계적 기법과 지식탐사기술(Knowledge Discovery Technology)을 이용하여 고객의 취향에 가장 부합하는 상품을 추천해주는 시스템으로서, 고객들의 편의를 도모하고 교차판매(Cross Sell) 및 매출 증대에 초점을 맞춘 시스템이다. 현재까지 추천시스템을 구현하기 위한 다양한 기법들이 개발되어 왔는데, 이 중에서 협업필터링(Collaborative Filtering) (Sarwar et al., 2000; Adomavicius and Tuzhilin, 2005; 김재경 등, 2005)이 가장 성공적인 추천 기법으로 알려져 있으며

Amazon.com, CDNow 등과 같은 대표적인 e-비즈니스 사이트에서 널리 활용되고 있다.

협업필터링은 상품을 추천하고자 하는 고객과 취향이 유사한 고객들의 의견을 반영하여 추천 대상 고객이 아직 구매하지 않은 상품에 대한 선호도를 예측한 후 선호도가 높을 것으로 예측되는 상품을 추천해주는 추천기법이다. 그러나 신상품의 경우에는 과거에 어떠한 고객도 구매한 적이 없어 이에 대한 선호도를 예측할 수 없으므로 추천이 불가능하게 되는 신상품 추천문제(Huang et al., 2004; Adomavicius and Tuzhilin, 2005)가 발생하게 된다. 예를 들어, 멀티미디어 사용자 환경

\* 본 논문은 2008년도 국민대학교 교내연구비를 지원받아 수행된 연구임.

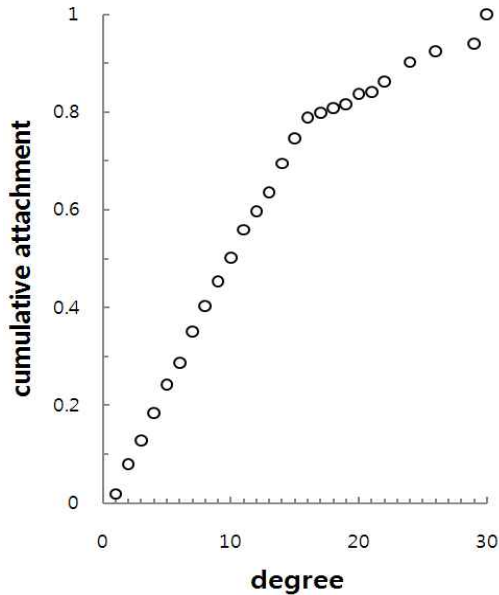
과 고객 취향의 급속한 변화에 따라 수시로 새로운 멀티미디어 콘텐츠가 제공되는 모바일 웹 사이트에 협업필터링을 적용할 경우, 기대하는 만큼 추천시스템의 효과를 얻기가 어렵다.

이와 같은 협업필터링의 한계점을 극복하기 위하여 많은 연구자들은 추천 대상 고객이 선호(또는 구매)하는 상품과 유사한 특성을 가진 상품을 추천하는 내용기반 필터링(Content Based Filtering)을 협업필터링과 결합하여 사용하였다(Melville, et al., 2001; Cho and Kim, 2004; Kim and Yum, 2005; 조운호 등, 2004; 김재경 등, 2005). 즉, 하이브리드 추천기법은 평가하지 않은 상품들의 평가 값을 내용기반 필터링으로 미리 예측함으로써 예측된 평가 값과 실제 사용자가 평가한 값들을 함께 고려하여 두 사용자의 유사 선호 경향을 분석한다. 따라서 두 사용자 간에 공통으로 평가한 상품이 없으면 두 사용자의 선호 경향이 유사한지 판별할 수 없게 되거나, 평가 결과가 없는 상품을 추천할 수 없는 협업필터링의 근본적인 문제를 해결함으로써 추천의 정확도를 높일 수 있고 신상품 추천도 가능하다. 현재까지의 개발된 하이브리드 추천기법은 정보검색(Information Retrieval) 분야의 연구결과를 바탕으로 영화(Schein et al., 2002; Celma et al., 2005), 음악(Kim et al., 2006), 웹페이지(Jian et al., 2004), 이미지(Kim, et al., 2004; Jang, et al., 2009) 추천 분야에서 키워드 또는 내용(content) 기반의 속성을 추출하고 활용하는 방식으로 연구가 진행되어 왔다. 그러나 전자제품, 의류, 가구 등 일반적인 상품의 경우에는 고객이 식별하는 속성이 상품별로 다양하고 달라서(예 : 가격, 브랜드, 스타일, 형태, 성능, 색상 등), 키워드나 내용을 추출하기가 어렵기 때문에 하이브리드 추천기법은 이들 상품을 추천하는데 한계가 있다. 따라서 상품 유형에 관계없이 고객에게 신상품을 효과적으로 추천

할 수 있는 새로운 접근방법이 모색되어야 한다.

최근에 사람 또는 사물 간의 관계를 네트워크 관점에서 분석할 수 있는 방법으로 사회연결망분석(Social Network Analysis)이 많이 연구되고 있다. 사회연결망분석은 개체의 상호작용에 관심을 두고, 개체간 연결 상태 및 연결 구조의 특성을 계량적으로 파악하여 시각적으로 표현하는 분석기법(Wasserman, 1994; 손동원, 2002; 김용학, 2003)으로 유전 네트워크(Kauffman, 1993), 교통 네트워크(윤성준, 2005), 조직 네트워크(최창현, 2006) 등의 구조 분석을 위해 널리 이용되고 있다. 최근에는 추천시스템 분야에서도 이를 활용하여 기존연구의 문제점을 근본적으로 해결하고자 하는 연구들이 진행 중이다(박종학 등, 2009).

일반적으로 사회연결망에 새로 추가되는 노드의 링크는 기존 노드와 무작위적으로 연결되는 것이 아니라 선호적 연결(Preferential attachment)을 하게 된다(Albert and Barabasi, 2002). 즉, 링크가 많은 노드에 새로운 노드가 연결될 가능성이 높다는 것이다. <그림 1>은 실제 구매 트랜잭션을 이용하여 고객들간의 구매 패턴이 유사한 상품들끼리 링크를 연결하는 상품 네트워크를 구성하고, 이 네트워크에 새로운 노드(신상품)들이 추가될 때 기존 노드(상품)의 연결정도(degree), 즉 연결된 링크 수가 어떻게 변화하는가를 도식한 것이다. 그림에서 보는 바와 같이 상품 네트워크 내의 모든 노드는 연결정도가 균일하게 증가하는 것이 아니라 기존 연결정도에 비례하여 링크 수가 증가함을 알 수 있다. 이는 인위적으로 관계를 설정한 상품 네트워크도 다른 사회연결망과 같이 선호적 연결을 하고 있음을 실증적으로 보여주고 있다. 이러한 선호적 연결의 의미는 많은 다른 상품들과 유사한 구매패턴을 보이는 상품일수록 신상품과도 구매관계가 연결될 확률이 높다는 것이고, 이를 고



<그림 1> 실제 상품 네트워크에서의 누적된 상호적 연결

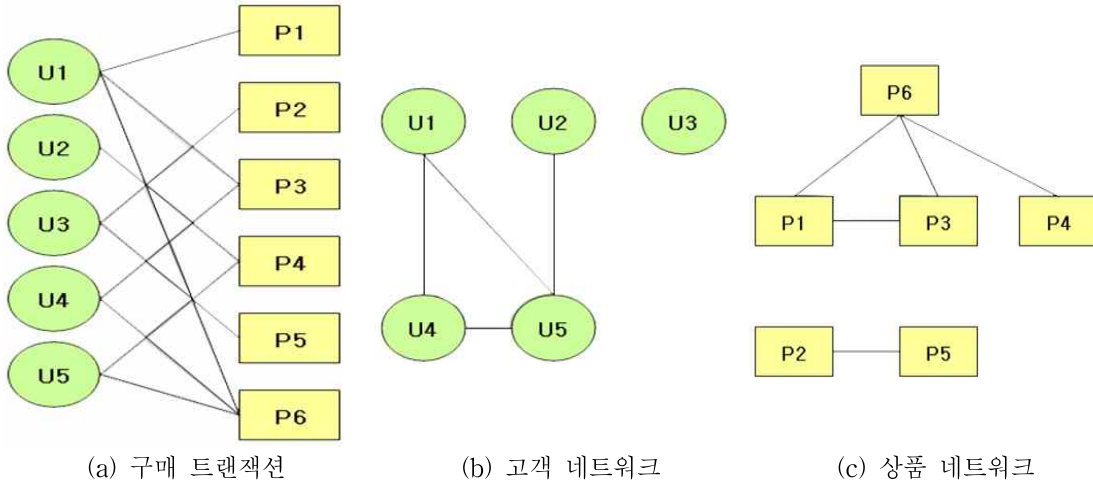
객 관점에서 바라보면 다수의 다른 상품들과 유사한 구매패턴을 보이는 상품들을 다양하게 구매했던 고객일수록 신상품도 구매할 가능성이 높다고 해석할 수 있다. 따라서 사회연결망분석을 이용하여 이러한 고객들을 식별하고 이들에게 신상품을 추천하면 무작위적인 추천보다 효과적이며, 상품의 유형에도 구매 받지 않는 신상품 추천이 가능하다.

이와 같은 아이디어를 기반으로 본 연구에서는 사회연결망분석에서 관계 및 구조적 특성을 분석하기 위해 널리 활용되고 있는 중심성(Centrality) 분석기법을 응용하여, 구매정보 외에 추가적인 정보를 요구하지 않으면서 모든 유형의 신상품을 적절한 고객에게 추천할 수 있는 새로운 방법을 제시한다. 또한 국내 유명 백화점의 구매 데이터를 이용하여 제시한 신상품 추천방법의 특성을 살펴보고 그 성능을 검증하고자 한다.

## 2. 사회연결망과 중심성분석

사회연결망은 Barnes(1954)에 의해 처음 사용된 용어로, 개인적인 인간관계가 확산되어 형성된 사람들 사이의 연결된 네트워크이다(손동원, 2002). 이러한 사회연결망은 사회과학뿐만 아니라 경영학, 응용과학 등 다양한 분야에서 응용되고 있으며, 분석 방법으로 사회연결망분석이 널리 이용되고 있다. 일반적으로 사회연결망분석에서는 분석하고자 하는 데이터를 행렬로 표현한다. 즉 행과 열이 만나는 셀에 특정 값을 입력하여 행과 열사이의 관계를 나타낸다. 여기서 행과 열에 같은 개체가 배열되는 것은 1원(1 Mode)자료라 하며, 다른 개체가 배열되는 경우를 2원(2 Mode) 자료라 한다. 또한 관계를 표현하는 기본적인 방법은 개체와 개체간의 관계가 존재하면 1, 존재하지 않으면 0으로 입력한다. 예를 들어 고객-상품 행렬 P에서 고객  $i$ 가 상품  $j$ 를 구매하면  $p_{ij} = 1$ , 구매하지 않으면  $p_{ij} = 0$ 으로 표시한다.

고객-상품 행렬에서 상품을 구매한 사람들 사이에 직접적인 상호작용의 관계가 없더라도, 관계를 인위적으로 설정하여 고객과 고객 또는 상품과 상품 사이의 관계를 나타낸 네트워크를 준 연결망(quasi network)이라 한다(김용학, 2003; 박종학 등, 2009). <그림 1>은 준 연결망의 예로, 구매 트랜잭션 (a)를 이용하여 (b), (c)와 같이 고객간의 관계 및 상품간의 관계를 표현하였다. 여기서 (b)를 고객 네트워크라 하며, 고객들이 서로 동일 상품을 1개 이상 구매하였다면 그 고객들은 직접적인 상호작용이 있다고 표현하였다. (c)는 상품 네트워크로서, 고객들이 특정 상품을 동시에 구매하면 그 상품은 서로 상호 관계가 있다고 표현하였다. 일반적으로 고객 또는 상품의 상호 관계를 파악하는 방법에는 동일 상품 구매 여부, 상관계수



<그림 2> 준 연결망의 예(박종학 등, 2009)

계산 등이 있다.

사회연결망분석은 여러 개의 기법들로 구성되어 있고 지금도 새로운 기법이 지속적으로 만들어지고 있다. 사회연결망분석에서 네트워크 구조를 파악하기 위한 기법으로 중심성(Centrality), 밀도(Density), 구조적 틈새(Structural hole), 집중도(Centralization) 등이 있다(Bonacich, 1987; 손동원, 2002; 김용학, 2003). 이 중에서 본 연구에서 활용하는 중심성은 한 행위자가 전체 네트워크에서 중심에 위치하는 정도를 표현하는 지표로 연결정도 중심성(Degree centrality), 근접 중심성(Closeness centrality), 매개 중심성(Betweenness centrality), 위세 중심성(Prestige centrality)이 있다(Bonacich, 1987; 손동원, 2002; 김용학, 2003).

- 연결정도 중심성 : 한 점에 직접적으로 연결되어 있는 점들의 합으로 얻어진다. 한 점에 얼마나 많은 다른 점들이 관계를 맺고 있는가를 통해서 그 점이 중심에 위치하는 정도를 계량화한 것이다. 연결정도 중심성은 한

점의 포인트 중심성을 측정하는 방법이다. 연결된 노드의 수가 많을수록 연결중심성이 높아진다.

- 근접 중심성 : 각 노드 간의 거리를 근거리하여 중심성을 측정하는 방법으로 연결정도 중심성과는 달리 직접적으로 연결된 노드뿐만 아니라 네트워크 내 간접적으로 연결된 모든 노드들 간의 거리를 합산하여 중심성을 측정한다는 특징을 가진다. 예를 들면 사람 A가 속해있는 어떤 집단에서 A가 집단 내에 있는 모든 사람들과 만난다고 가정하자. A는 집단 내에 있는 모든 사람들을 알지는 못한다. 하지만 A가 아는 사람들을 거쳐서 가면 모든 사람들을 알 수 있게 된다. 이때 가장 짧은 경로로 모든 사람을 알게 되는 것이 근접중심성이다. 근접중심성은 각 노드 간의 거리를 근거리하여 중심성을 측정하는 지표로 한 노드로부터 다른 노드에 도달하려면 필요한 최소 단계의 합으로 정의한다. 근접중심성이 높을수록 네트워크의 중앙에 위치하게 된다.

- 매개 중심성 : 네트워크 내에서 한 점이 담당하는 매개자 혹은 중재자 역할의 정도로써 중심성을 측정하는 방법이다. 예로 A, B, C 세 사람이 있다. 이 중 A와 B는 C를 통해서만 관계를 맺을 수 있는 경우 C는 잠재적으로 다른 사람들 사이를 통제할 수 있는 ‘브로커’ 또는 ‘문지기’의 역할을 한다. 이 경우 C는 매개중심성이 높다. 한 노드가 연결망 내의 다른 노드들 사이의 최다 경로 위에 위치하면 할수록 그 노드의 매개중심성이 높아진다. 매개중심성 또한 네트워크간 비교를 위해 상대적 매개중심성이 사용된다.
- 위세 중심성 : 연결된 노드의 중요성에 가중치를 주어 노드의 중심성을 측정하는 방법이다. 즉, 위세가 높은 사람들과 관계가 많을수록 자신의 위세가 높아진다는 개념이다. 예를 들면 영향력이 높은 사람과의 단 하나의 연결이 그렇지 않은 다른 여러 사람들과 관계를 맺고 있는 경우보다 자신의 영향력(위세)을 증가시킬 수 있다. 위세 중심성의 일반적인 형태는 보나시치(Bonacich) 권력지수로 불리며, 자신의 연결정도 중심성으로부터 발생하는 영향력과 자신과 연결된 타인의 영향력을 합하여

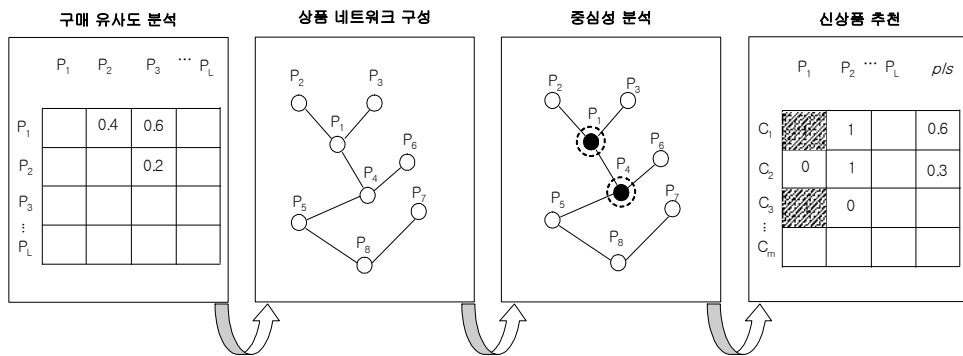
위세 중심성을 결정한다.

위의 네 가지 중심성은 각각 서로 다른 특성을 갖고 있고, 추천시스템 분야에 있어 현재까지 이들을 활용하는 연구가 미진하기 때문에 어느 것이 신상품 추천에 더 적합한지 이론적으로 규명하기 어렵다. 따라서 본 연구에서는 실제 데이터를 가지고 위의 네 가지 중심성 분석기법을 신상품 추천에 각각 적용한 후 추천 성능의 차이를 실증적으로 검증하고자 한다.

### 3. 사회연결망분석 기반 신상품 추천 방법

#### 3.1 신상품 추천 프로세스

<그림 3>은 사회연결망분석을 활용한 신상품 추천 방법의 개괄적인 절차를 보여준다. 총 4단계로 추천 프로세스가 구성되는데, 첫 번째는 ‘구매 유사도 분석’ 단계로 고객들의 과거 구매 데이터를 가지고 모든 기존 상품들에 대하여 서로간의 구매 패턴에 대한 상관관계를 계산하여 상품 간 구매 유사도를 구한다. 두 번째는 ‘상품 네트워크 구성’ 단계로 고객들 간의 구매 패턴이 유사한 상품들끼



<그림 3> 신상품 추천 프로세스

리 서로 링크로 연결하여 하나의 사회연결망을 만드는 단계이다. 세 번째는 ‘중심성 분석’ 단계로 사회연결망에서 각 노드(상품) 별로 중심성을 측정하여 중심성이 높은 상품들을 찾는 단계이다. 마지막 ‘신상품 추천’ 단계에서는 중심성이 높은 상품들을 다양하게 구매했던 고객들을 찾고 이들에게 신상품을 추천한다.

### 3.2 단계 1 : 구매 유사도 분석

단계1에서는 기존 상품들에 대한 고객들의 구매성향을 분석하고 상품 간 유사도를 구하는 단계인데, 구매성향을 분석하는 과정과 유사도를 구하는 과정으로 나누어진다. 구매성향을 분석하기 위해서는 현재 데이터베이스 내에 있는 기존  $M$ 명의 고객이  $L$ 개의 상품 중 어떤 것을 구매했는지를 알아본다. 특정 고객이 특정 상품을 한 개 이상 구매한 경우를 1, 그렇지 않은 경우를 0으로 설정하면, 고객-상품 행렬  $\mathbf{P} = (p_{ij})$ 는 식 (1)과 같이 표현될 수 있다. 여기서  $P$ 는 상품과 상품 사이에는 아무런 직접적인 관계가 존재하지 않지만 상품을 구매한 고객들에 의하여 관계가 인위적으로 설정되는 준 연결망을 분석하기 위해 사용된다.

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

다음 과정으로, 피어슨 상관계수(Pearson Correlation Coefficient)를 사용하여 식 (2)과 같이 두 상품  $a$ 와  $b$ 의 구매 유사도  $sim(a, b)$ 를 구한다. 식 (2)에서  $p_{ka}$ 와  $p_{kb}$ 는 상품  $a$ 와  $b$ 에 대한 고객  $k$ 에 대한 구매여부를 나타내며,  $\bar{p}_a$ ,  $\bar{p}_b$ 는 상품  $a$ 와  $b$ 의 구매여부 평균값을 의미한다. 피어슨 상관계수를 이용하여 계산한 상품 간 유사도는 -1에서 +1 사

이의 값을 갖는다. 유사도 값이 ‘1’일 경우는 같은 고객들이 두 상품을 구매했다는 것을 의미하며, ‘-1’인 경우는 두 상품을 구매한 고객들이 서로 다르다는 것을 의미한다.



(2)

### 3.3 단계 2 : 상품 네트워크 구성

단계 2에서는 단계 1에서 계산한 상품 간 구매 유사도로부터 구매 패턴이 유사한 상품들을 서로 연결하는 상품 네트워크를 구성하는 단계이다. 모든 상품에 대하여 임의의 두 상품 간 구매 유사도가 특정 임계치  $p$  이상인 경우 이들을 링크로 연결한다. 즉, 정해진 수준 이상으로 구매패턴에 있어서 상관관계가 있어야 사회적 관계가 설정된다는 의미이다. 상품  $a$ 와  $b$ 간의 연결여부  $link(a, b)$ 는 식 (3)과 같이 정의되며,  $link(a, b)$ 가 1인 경우  $a$ 와  $b$ 간에 링크가 형성된다.

$$link(a, b) = \begin{cases} 1, & sim(a, b) \geq p \\ 0, & sim(a, b) < p \end{cases} \quad (3)$$

이와 같이 형성된 상품 네트워크에서는 비슷한 구매패턴을 가진 상품이 이웃 노드로 연결된다. 한 상품에게 연결된 링크가 많다는 것은 이웃 노드가 많다는 것이고, 이웃 노드가 많다는 것은 그 상품이 다른 많은 상품들과 고객들 간의 유사한 구매 패턴을 보인다는 것을 의미한다.

### 3.4 단계 3 : 중심성 분석

단계 3에서는 상품 네트워크에서 중심성 값이

높은 상품을 찾는다. 즉, 전체 네트워크에서 각 상품의 중심성을 계산하여 중심성 값이 높은 상위  $K$ 개의 상품(중심상품)을 뽑고, 신상품 추천 대상 고객을 선정하기 위해 이들을 이용하는 것이다. 제2절에서 논의한 바와 같이 중심성 측정지표에는 ‘연결정도 중심성’, ‘근접 중심성’, ‘매개 중심성’, ‘위세 중심성’ 등 4가지가 있는데 이중에 하나를 선택하여 각 상품의 중심성을 측정한다. 임의의 상품  $a$ 의 중심성  $cen(a)$ 은 각 중심성 측정지표에 따라 아래와 같이 계산된다.

- 연결정도 중심성 측정 : 식 (3)과 같이 두 노드  $a$ 와  $j$ 의 연결여부를  $link(a, j)$ 라고 할 때, 상품  $a$ 의 연결정도 중심성  $cen(a)$ 는 다음과 같이 계산된다.

$$cen(a) = \sum_{j=1}^L link(a, j) \quad (4)$$

- 근접 중심성 측정 : 두 노드  $a$ 와  $j$ 를 잇는 최단 경로의 거리를  $dist(a, j)$ 라고 할 때, 상품  $a$ 의 근접 중심성  $cen(a)$ 는 다음과 같이 계산된다.

$$cen(a) = \sqrt{\sum_{j=1}^L dist(a, j)} \quad (5)$$

- 매개 중심성 측정: 두 노드  $a$ 와  $j$  사이에 존재하는 최단경로의 경우 수를  $geod(j, k)$ 라 하고, 두 노드  $j$ 와  $k$  사이에 존재하는 노드  $a$ 를 경유하는 경우의 수를  $geod(j, k, a)$ 라고 하면, 상품  $a$ 의 매개 중심성  $cen(a)$ 는 다음과 같이 계산된다.

$$cen(a) = \sum_{j,k} geod(j, k, a) / geod(j, k) \quad (6)$$

- 위세 중심성 측정 : 상품  $a$ 의 위세 중심성  $cen(a)$ 는 다음과 같이 계산된다.

$$cen(a) = \sum_{i=1}^L cen(a) \times link(a, j) \quad (7)$$

### 3.5 단계 4 : 신상품 추천

단계 4는 단계 3에서 추출한  $K$ 개의 중심상품을 구매한 고객들을 찾은 후 각 고객의 구매가능성을 점수화하여, 구매가능성 점수가 높은 상위 고객을 신상품 추천 대상 고객으로 선정하는 단계이다. 임의의 고객  $c$ 의 신상품에 대한 구매가능성 점수(purchase likelihood score)  $pls(c)$ 는 식 (8)과 같이 표현된다.

$$pls(c) = \frac{\sum_{j \in C} p_{cj} \times cen(j)}{\sum_{j \in C} cen(j)} \quad (8)$$

식 (8)에서  $p_{cj}$ 는 고객  $c$ 의 상품  $j$ 에 대한 구매여부를 나타내며,  $cen(j)$ 는 상품  $j$ 의 중심성값,  $C$ 는 단계 3에서 선택된  $K$ 개의 중심상품을 의미한다.

$K$ 개의 중심상품을 구매한 고객들 각각에 대하여  $pls$  값을 계산한 후 값이 큰 순으로 정렬하여 상위  $N$ 명의 고객에게 신상품을 추천한다.

## 4. 신상품 추천 적용사례

이 절에서는 제 3장에서 설명한 1단계에서 4단계까지의 신상품 추천 프로세스를 구체적인 사례를 통해 설명하기로 한다. 이 사례에서는 임의의 인터넷 상점에서 10개의 상품을 팔고 있다고 가정한다. 이 인터넷 상점에서 상품을 구매하였던 고객은 총 10명이며, 이들은 서로 아무 연관성도 없고 자신

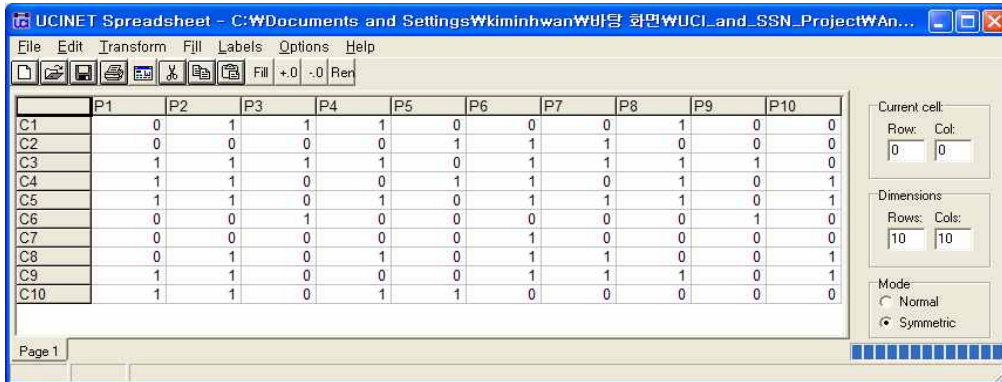
<표 1> 구매 트랜잭션

고객	상품	상품구매개수
C <sub>1</sub>	P <sub>2</sub> , P <sub>3</sub> , P <sub>4</sub> , P <sub>8</sub>	4
C <sub>2</sub>	P <sub>5</sub> , P <sub>6</sub> , P <sub>7</sub>	3
C <sub>3</sub>	P <sub>1</sub> , P <sub>2</sub> , P <sub>3</sub> , P <sub>4</sub> , P <sub>6</sub> , P <sub>7</sub> , P <sub>8</sub> , P <sub>9</sub>	8
C <sub>4</sub>	P <sub>1</sub> , P <sub>2</sub> , P <sub>5</sub> , P <sub>6</sub> , P <sub>8</sub> , P <sub>10</sub>	6
C <sub>5</sub>	P <sub>1</sub> , P <sub>2</sub> , P <sub>4</sub> , P <sub>6</sub> , P <sub>7</sub> , P <sub>8</sub> , P <sub>10</sub>	7
C <sub>6</sub>	P <sub>3</sub> , P <sub>9</sub>	2
C <sub>7</sub>	P <sub>6</sub>	1
C <sub>8</sub>	P <sub>2</sub> , P <sub>4</sub> , P <sub>6</sub> , P <sub>7</sub> , P <sub>10</sub>	5
C <sub>9</sub>	P <sub>1</sub> , P <sub>2</sub> , P <sub>6</sub> , P <sub>7</sub> , P <sub>8</sub> , P <sub>10</sub>	6
C <sub>10</sub>	P <sub>1</sub> , P <sub>2</sub> , P <sub>4</sub> , P <sub>5</sub>	4

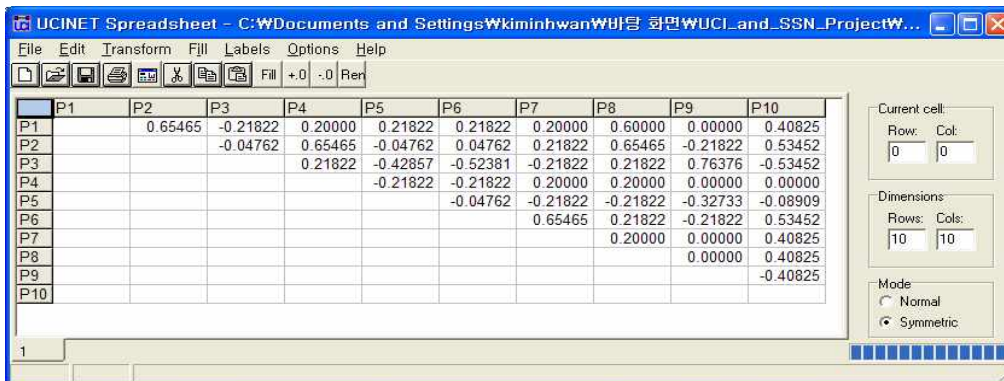
외 다른 구매자가 누구인지 모르는 상황이라고 가정한다. 또한 본 연구에서 제시하는 방법에 따라 사회연결망 분석도구인 UCINET 6.0을 사용하여 추천시스템을 구축하였고 D일부터 신규 고객에 대한 추천을 시작한다고 하자.

#### 4.1 구매 유사도 분석

10명의 기존고객  $C = \{C_1, C_2, \dots, C_{10}\}$ 이 10개의 상품  $P = \{P_1, P_2, \dots, P_{10}\}$ 을 D-1일까지 구매한 이력을 정리하면 <표 1>과 같다고 가정하자. 10명의 고객들로부터 가장 많이 구매된 상품은 P<sub>2</sub>, P<sub>6</sub>이

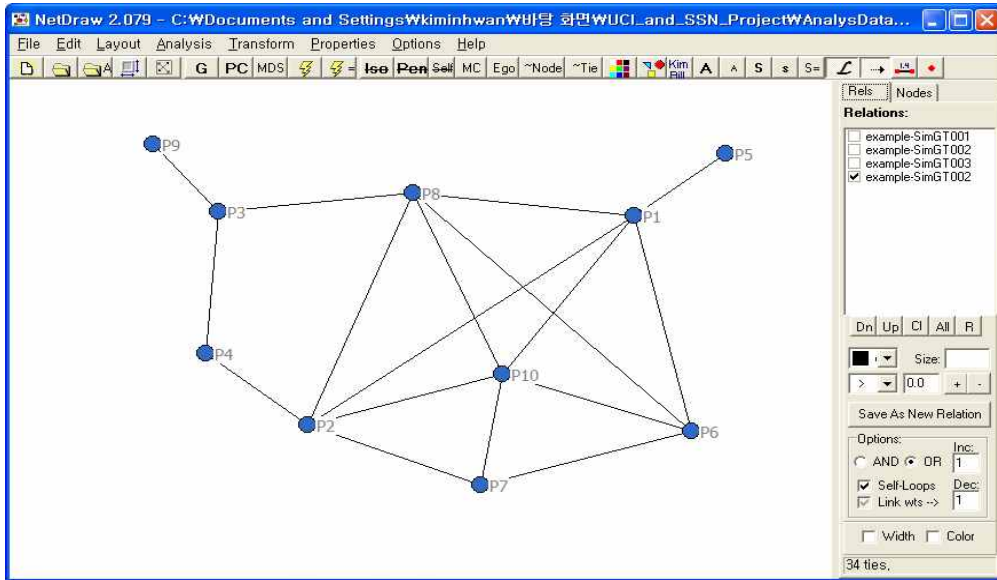


<그림 4> 고객-상품 행렬



<그림 5> 상품간 구매 유사도





<그림 6> 상품 네트워크

고, 가장 적게 팔린 상품은 P<sub>9</sub>이다.

<표 1>을 식 (1)을 대입하여 구매는 '1', 비구매는 '0'으로 나타내면 <그림 4>와 같은 고객-상품 행렬을 만들 수 있다. 본 사례에서는 없지만 만약 한 고객이 같은 물건을 2개 이상 구매하였을 경우에도 '1'로 표시한다.

<그림 4>의 고객-상품 행렬에 식 (2)를 적용하여 10개 상품의 상호간 유사도(상관관계)를 구하면 <그림 5>와 같은 결과를 얻는다. 유사도가 높다는 것은 두 상품을 같이 구매하거나 둘 다 구매하지 않은 고객들이 많다는 것을 의미한다.

가장 높은 유사도를 보인 상품은 유사도가 0.764인 P<sub>3</sub>와 P<sub>9</sub>이다. P<sub>3</sub>를 구매한 고객은 C<sub>1</sub>, C<sub>3</sub>, C<sub>6</sub> 고객이고 P<sub>9</sub>를 구매한 고객은 C<sub>3</sub>, C<sub>6</sub> 고객으로서 C<sub>1</sub> 고객의 구매여부만 다르다. 가장 낮은 유사도를 보인 상품은 유사도가 -0.535인 P<sub>3</sub>와 P<sub>10</sub>로서, 두 상품을 같이 구매한 고객은 한 명도 없다.

#### 4.2 상품 네트워크 구성

이전 단계에서 계산된 유사도 값이 임계치  $\rho$ 보다 크면 두 상품의 관계를 링크로 연결하고 유사도가  $\rho$ 보다 낮으면 두 상품의 관계를 형성하지 않는다.  $\rho$ 값은 적용분야 및 데이터의 특성에 따라 달라지는데, 추천시스템 관리자가 그 값을 적절히 선택한다. 이 사례에서  $\rho$ 값을 0.2로 설정하면, <그림 6>과 같은 상품 네트워크가 만들어진다. 이 네트워크에서 특징적인 것은 가장 많이 팔린 상품이 반드시 가장 많은 링크, 즉 연결관계를 갖지는 않는다는 것이다. 예를 들어, 7개가 팔린 상품 P<sub>6</sub>의 링크 수는 4인 반면, 4개가 팔린 상품 P<sub>5</sub>의 링크 수는 5임을 알 수 있다. 이것은 관계를 형성할 때 많이 팔린 상품이 많은 상품과 이웃하는 것이 아니고, 여러 고객들이 공통으로 많이 구매한 상품이 많은 다른 상품과 이웃으로 연결된다는 것을 의미한다.

	1	2	3
	Degree	NrmDegree	Share
1 P1	5.000	55.556	0.147
2 P2	5.000	55.556	0.147
8 P8	5.000	55.556	0.147
10 P10	5.000	55.556	0.147
6 P6	4.000	44.444	0.118
3 P3	3.000	33.333	0.088
7 P7	3.000	33.333	0.088
4 P4	2.000	22.222	0.059
9 P9	1.000	11.111	0.029
5 P5	1.000	11.111	0.029

(a) 연결정도 중심성

	1	2
	Farness	nCloseness
8 P8	13.000	69.231
2 P2	14.000	64.286
1 P1	14.000	64.286
10 P10	14.000	64.286
6 P6	16.000	56.250
3 P3	17.000	52.941
4 P4	18.000	50.000
7 P7	19.000	47.368
5 P5	22.000	40.909
9 P9	25.000	36.000

(b) 근접 중심성

	1	2
	Betweenness	nBetweenness
8 P8	11.150	30.972
3 P3	8.700	24.167
1 P1	8.450	23.472
2 P2	7.300	20.278
10 P10	1.950	5.417
4 P4	1.500	4.167
6 P6	1.500	4.167
7 P7	0.450	1.250
9 P9	0.000	0.000
5 P5	0.000	0.000

(c) 매개 중심성

	1	2
	Eigenvec	nEigenvec
1 P1	0.417	58.959
2 P2	0.407	57.532
3 P3	0.139	19.651
4 P4	0.129	18.278
5 P5	0.099	13.963
6 P6	0.376	53.203
7 P7	0.293	41.432
8 P8	0.425	60.048
9 P9	0.033	4.654
10 P10	0.454	64.219

(d) 위세 중심성

<그림 7> 중심성 분석 결과

### 4.3 중심성 분석

3.4절의 4가지 중심성 측정 지표에 따라 모든 상품에 대해 중심성을 계산하면 <그림 7>과 같은 결과를 얻는다. 그림으로부터 각 지표에 따라 매우 상이한 분석결과를 보여주고 있음을 알 수 있다. 추천시스템 관리자가 중심성 측정 지표로 ‘연결정도 중심성’을 선택하고 중심상품의 수  $K$ 를 4개로 설정했다고 가정하자. 그러면, <그림 7(a)>의 분석 결과로부터 연결정도가 5로 가장 높은 4개의 상품 {P1, P2, P8, P10}을 중심상품으로 선정하게 된다.

### 4.4 신상품 추천

마지막 단계에서는 중심상품을 적어도 한 개 이상 구매한 고객 중에서  $p_{ls}$  값이 높은  $N$ 명의 고객을 대상으로 신상품을 추천한다. 먼저, 식 (8)을 이용하여 각 고객의  $p_{ls}$  값을 구하면 <표 2>와 같은 결과를 얻게 된다.

여기서  $C_2, C_6, C_7$  고객은 중심상품을 하나도 구매하지 않았기 때문에 추천 대상에서 제외되었다. 추천 대상 고객 수  $N$ 을 3명으로 설정한다고 하면,  $p_{ls}$  값이 가장 높은 3명의 고객은 {C4, C5, C9}이고,

<표 2> 고객별  $\rho/s$  값

	C <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>	C <sub>4</sub>	C <sub>5</sub>	C <sub>6</sub>	C <sub>7</sub>	C <sub>8</sub>	C <sub>9</sub>	C <sub>10</sub>
$\rho/s$	0.5	-	0.75	1.0	1.0	-	-	0.5	1.0	0.5

따라서 이들에게  $D$ 일 이후부터 신상품을 추천하게 된다.

## 5. 성능 평가

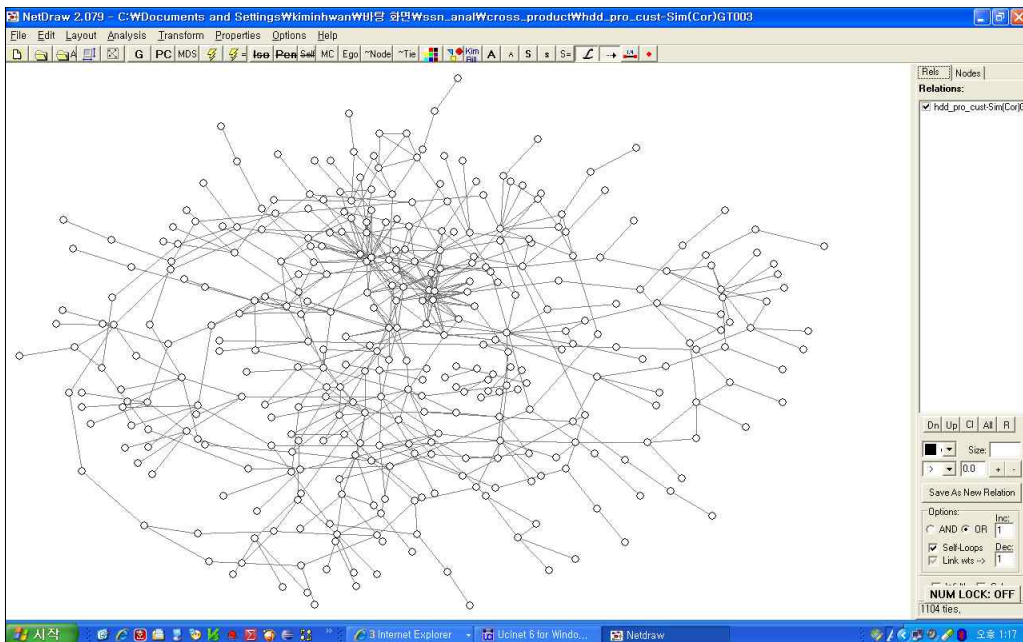
### 5.1 실험 데이터 및 방법

본 연구에서 제시한 신규고객 추천방법의 성능을 평가하기 위하여 국내 유명 백화점 중의 하나인 H백화점의 구매 트랜잭션 데이터를 사용하여 실험하였다. 실험 도구로 사회연결망 분석을 위해 UCINET 6.0을, 자료처리와 결과분석을 위해 SPSS Clementine 11.1을 각각 사용하였으며, 구체적인

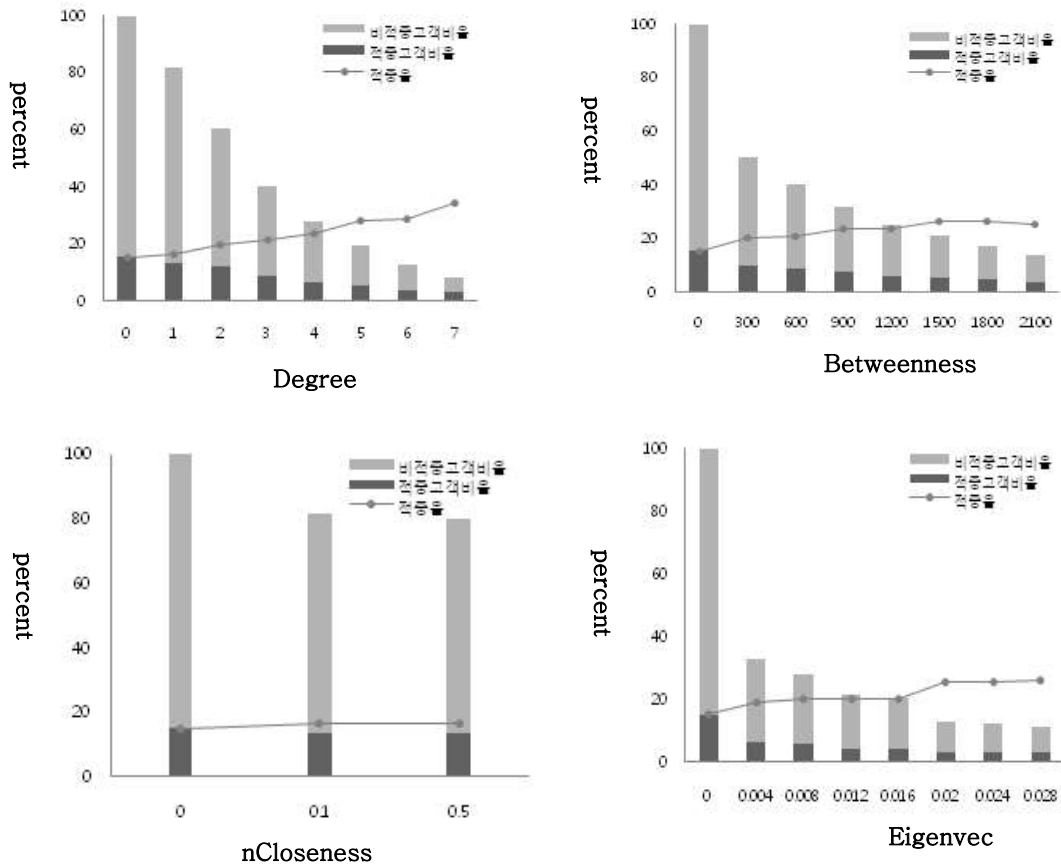
실험 방법은 다음과 같다.

1) 2000년 5월 1일부터 2001년 4월 30일까지의 12개월 데이터(총 구매건수 10,734건)를 각 6개월씩 학습기간(2000. 5. 1~2000. 10. 31)과 적용기간(2000. 11. 1~2001. 4. 30)으로 분할하였다. 학습기간 동안의 트랜잭션에 나타나는 고객은 678명이고 상품은 524개였다.

2) 678×524 고객-상품 행렬 데이터로부터 각 상품의 유사도를 계산하고 임계치  $\rho$ 를 0.3으로 설정하여 <그림 8>과 같은 상품 네트워크를 얻었다. 이 과정에서 다른 상품과의 구매 상관관계가 약해 링크가 한 개도 생성되지 않은 174개의 상품은 네트워크에서 배제되었고, 나머지 350개의 상품만으



<그림 8> H백화점 구매 트랜잭션으로부터 생성된 상품 네트워크



<그림 9> 중심성 분석기법 별 적중률

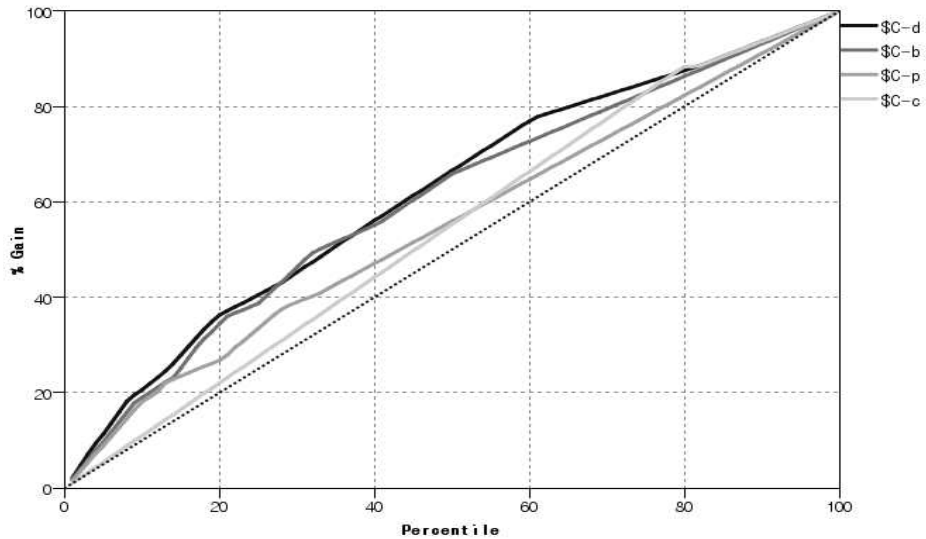
로 최종 네트워크가 구성되었다.

3) 상품 네트워크의 모든 노드에 대해 4가지 종류의 중심성을 각각 측정하고 종류별로 중심성이 높은 순으로 상품을 정렬하였다.

4) 적용기간 동안 구매가 발생했던 상품 가운데 학습기간에 구매가 한 번도 일어나지 않았던 상품을 신상품으로 정의하였는데, 이에 따라 총 20개의 상품이 신상품으로 식별되었다. 이들 신상품을 구매한 총 120명의 고객 중에서 학습기간에도 구매 기록이 있는 103명을 신상품 구매고객으로 선정하

였다.

5) 각 중심성 지표 별로, 중심상품 수  $K$ 를 1에서 가능한 값(이 실험에서 근접 중앙성의 경우는 0.5, 0.1, 0 등 3가지 값만 측정됨)까지 변화시켜가면서, 학습기간에  $K$ 개의 중심상품을 구매한 고객 중에 몇 명이 신상품 구매고객이 되었는지, 즉 적중률(hit ratio)을 측정하였다. 분석을 명료화하기 위하여, 본 실험에서는  $pls$ 를 적용하지 않고 중심상품을 구매한 모든 고객들에 대해 신상품을 추천한다고 가정하였다.



<그림 10> 중심성 분석기법 별 이득 도표

## 5.2 실험 결과 및 분석

중심성 지표, 중심상품 수를 변화시켜 가면서 얻은 실험결과를 도식하면 <그림 9>와 같다. 그림에서 보면, 중심성 측정 지표에 따라 적중률에 있어 분명한 차이가 있다는 것을 알 수 있다. 적중률 최대치를 기준으로 할 때, ‘연결정도 중심성’이 약 34.5%로 가장 높고, ‘매개 중심성’과 ‘위세 중심성’이 각각 26.5%와 25.9%로 거의 비슷하며, ‘근접 중심성’이 16.8%로 가장 낮다. 적중률 34.5%의 의미는 3명에게 신상품을 추천했을 때 그 중 한 명은 신상품을 구입한다는 것으로 상당히 높은 추천 성능을 나타낸다. 또한 ‘연결정도 중심성’은 다른 중심성 지표에 비해 중심상품의 수에 따라 적중률이 많이 변화하고 있음을 보여주고 있다. 이는 ‘연결정도 중심성’이 다른 지표에 비해 측정값(degree)이 연속적이고 이에 따른 노드의 분포도 어느 정도 산재해 있기 때문으로 판단된다. 극단적으로 ‘근접 중심성’은 측정값(nCloseness)이 적고 비연속적이어서 중심상품의 수에 따른 적중률에 차이

가 없다는 것이 이를 입증한다고 볼 수 있다.

<그림 10>은 각 중심성 지표 별로 얻은 적중률을 토대로 이득도표(Gain chart)<sup>1)</sup>를 도식한 것이다. 이득도표는 좌표 (0, 0)에서 좌표 (100, 100)으로 연결되는 증가함수 형태를 취하게 되는데, 추천 기법이 완전히 무의미한 경우(예 : 무작위적 추천)는  $Y = X$ 인 직선이 되고 적중률이 높으면 높을수록 이 직선에서 위로 멀어지게 된다. 이득도표에서 볼 수 있듯이 ‘연결정도 중심성’(\$C-d)의 경우 20% 선에서의 이득이 대략 37%인 것을 볼 수 있다. 이것의 의미는 제시한 추천방법을 토대로 고객을 정렬하여 상위 20%를 신상품 추천 대상자로 예측하면, 전체 신상품 구매자의 37%가 이중에 속한다는 것을 나타낸다. ‘매개 중심성’(\$C-b)의 경우는 20% 선에서의 이득이 대략 34%로 ‘연결정도 중심성’과 거의 차이가 없지만, ‘위세 중심성’(\$C-p)과 ‘근접

1) X축에는 0에서 100%까지 퍼센트 값을, Y축에는 그 퍼센트에 들어있는 hit 수/전체 hit 수\* 100(%)를 계산하여 도식한 그래프(허명희 등, 2008).

중심성'(\$C-c)은 이득이 크게 의미가 없는 것으로 나타났다. 위와 같은 실험결과를 종합해 볼 때, 적절한 중심상품 수를 설정하고 '연결정도 중심성'을 중심성 분석기법으로 활용할 경우 본 연구에서 제시하는 추천방법은 성능(적중률)과 ROI(이득) 측면에서 모두 우수한 신상품 추천방법이라고 판단된다.

## 6. 결론

본 연구에서는 사회연결망분석에서 관계·구조적 특성을 분석하기 위해 널리 활용되고 있는 중심성 개념을 적용하여 상품으로 구성된 네트워크에서 중심이 되는 상품을 찾고 이를 기반으로 신상품을 구매할 가능성이 높은 고객을 대상으로 신상품을 추천하는 방법을 제시하였다. 실험 결과, 제시한 추천방법은 기존의 신상품 추천방법들과 달리 구매정보 외에 추가 정보가 필요하지 않으며, 모든 종류의 신상품을 적절한 고객에게 추천할 수 있는 새로운 방법임이 입증되었다.

본 연구의 한계 및 추후 연구과제로는 다음과 같은 것들이 있다. 첫째, 이 연구에서는 과거 구매 데이터를 이용하여 제시한 방법의 추천 적중률을 측정함으로써 그 성능을 평가하였다. 실제 고객들은 보다 다양한 상황에서 상품에 대한 구매결정을 내리므로 향후 연구에서는 실제 고객이 직접 참여하는 실험을 통해 제시한 추천방법을 검증할 필요가 있다. 둘째, 제시하는 추천방법에서 중심상품 분석 시 한가지 측정지표만을 사용하고 있지만, 모든 지표를 함께 고려하는 방안을 만들어 중심성을 다각도로 분석한다면 보다 향상된 결과를 얻을 수 것으로 예상된다. 셋째, 본 연구에서는 사회연결망 분석 기법 중 중심성 개념만을 사용하는 추천방법

을 제시하였다. 하지만, 밀도, 집중도, 구조적 등위성 등 사회연결망의 전반적 성격을 규명하는 분석 기법을 활용하여 상품 네트워크의 성격을 규정하고 상품 네트워크의 성격에 따라 신상품 추천방법을 다르게 적용하는 것이 보다 실제적일 것이다.

## 참고문헌

- 김용학, *사회연결망 분석*, 박영사, 2003.
- 김재경, 조운호, 김승태, 김혜경, “모바일 전자상거래 환경에 적합한 개인화된 추천시스템”, *경영정보학연구*, 15권 3호(2005), 223~241.
- 박종학, 조운호, 김재경, “사회연결망 : 신규고객 추천문제의 새로운 접근법”, *지능정보연구*, 15권 1호(2009), 123~140.
- 손동원, *사회 네트워크 분석*, 경문사, 2002.
- 윤성준, “데이터마이닝 기법을 통한 백화점의 고객이탈예측 모형 연구”, *한국마케팅저널*, 6권 4호(2005), 45~72.
- 조운호, 박수경, 안도현, 김재경, “재구성된 제품 계층도를 이용한 협업 추천 방법론 및 그 평가”, *한국경영과학회지*, 29권 2호(2004), 61~77.
- 최창현, “조직의 비공식 연결망에 관한 연구: 사회연결망 분석의 적용”, *한국사회와행정연구*, 17권 1호(2006), 1~23.
- 허명희, 이용구, *데이터마이닝 모델링과 사례(제2판)*, 한나래출판사, 2008.
- Adomavicius, G. and A. Tuzhilin, “Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions”, *IEEE Transactions on Knowledge and Data Engineering*, Vol.17, No.6 (2005), 734~749.
- Albert, R. and A. L. Barabasi, “Statistical me-

- chanics of complex networks”, *Review of Modern Physics*, Vol.74(2002), 47~97.
- Barnes, J., “Class and committees in a Norwegian island parish”, *Human Relations*, Vol.7(1954), 39~58.
- Bonacich, P., “Power and Centrality : A Family of Measures”, *American Journal of Sociology* 92, 1987, 1170~1182.
- Celma, O., M. Ramirez, and P. Herrera, “Foafing the music : A music recommendation system based on RSS feeds and user preferences”, *ISMIR*, 2005, 464~457.
- Cho, Y. H. and J. K. Kim, “Application of Web usage mining and product taxonomy to collaborative recommendations in e-commerce”, *Expert Systems with Applications*, Vol.26, No.3(2004), 234~246.
- Huang, Z., H. Chen, and D. Zeng, “Applying Associative Retrieval Techniques to Alleviate the Sparsity Problem in Collaborative Filtering”, *ACM Transactions on Information Systems*, Vol.22, No.1(2004), 116~142.
- Jang, M. K., J. K. Kim, H. K. Kim, and Y. H. Cho, “A New Item Recommendation Procedure Using Preference Boundary”, *APJIS*, Vol.19, No.4(2009).
- Jian, C. and Y. Jian, H. Jin, “Recommendation of New Items Based on Indexing Techniques”, *Proceedings of the Third International Conference on Machine Learning and Cybernetics*, 2004, 1168~1172.
- Kauffman, S., *The Origins of Order*, Oxford University Press, 1993.
- Kim, C. Y., J. K. Lee, Y. H. Cho, and D. H. Kim, “VISCORS : A Visual-Content Recommender for the Mobile Web”, *IEEE Intelligent Systems*, Vol.19, No.6(2004), 32~39.
- Kim, D. and B. J. Yum, “Collabrative Filtering based on Iterative Principal Component Analysis”, *Expert System with Applications*, Vol.28(2005), 823~830.
- Kim, J. K., Y. H. Cho, M. Y. Kang, and H. K. Kim, “A Hybrid Multimedia Contents Recommendation Procedure for a New Item Problem in M-commerce”, *Journal of Korea Intelligent Information Systems Society*, Vol. 12, No.2(2006), 1~15.
- Melville, P., R. J. Mooney, and R. Nagarajan, “Content-boosted Collaborative Filtering”, *Proceeding SIGIR 2001 Workshop on Recommender Systems*, 2001.
- Park, S. T., D. Pennock, O. Madani, N. Good, and D. DeCoste, “Naive Filterbots for Robust Cold-Start Recommendations”, *KDD*, 2006.
- Sarwar, B., G. Karypis, J. A. Konstan, and J. Riedl, “Analysis of recommendation algorithms for e-commerce”, *Proceedings of ACM E-commerce 2000 conference*, 2000, 158~167.
- Schein, A. I., A. Popescul, D. M. Pennock, and L. H. Ungar, “Methods and Metrics for Cold-Start Recommendations”, *SIGIR*, 2002.
- Wasserman, S., *Social network analysis : methods and applications*, Cambridge University Press, 1994.





Abstract

## Social Network Analysis for New Product Recommendation

Yoonho Cho\* · Jounghae Bang\*

Collaborative Filtering is one of the most used recommender systems. However, basically it cannot be used to recommend new products to customers because it finds products only based on the purchasing history of each customer. In order to cope with this shortcoming, many researchers have proposed the hybrid recommender system, which is a combination of collaborative filtering and content-based filtering. Content-based filtering recommends the products whose attributes are similar to those of the products that the target customers prefer. However, the hybrid method is used only for the limited categories of products such as music and movie, which are the products whose attributes are easily extracted. Therefore it is essential to find a more effective approach to recommend to customers new products in any category.

In this study, we propose a new recommendation method which applies centrality concept widely used to analyze the relational and structural characteristics in social network analysis. The new products are recommended to the customers who are highly likely to buy the products, based on the analysis of the relationships among products by using centrality. The recommendation process consists of following four steps; purchase similarity analysis, product network construction, centrality analysis, and new product recommendation. In order to evaluate the performance of this proposed method, sales data from H department store, one of the well-known department stores in Korea, is used.

**Key Words** : New Product Recommendation, Social Network Analysis, Product Network, Centrality Analysis, Collaborative Filtering

---

\* School of Business Administration, Kookmin University

## 저자 소개



조운호

현재 국민대학교 경영학부 e-비즈니스전공 부교수로 재직 중이다. 서울대학교 계산통계학과(전산학전공)를 졸업하고, KAIST 경영정보공학과에서 석사, KAIST 경영공학과에서 박사학위를 취득하였으며, LG전자(주)에서 6년간 주임연구원으로 재직하였다. 주 연구분야는 추천시스템, 모바일비즈니스, 고객관계관리, 데이터마케팅 등이다.



방정혜

현재 국민대학교 경영학부 마케팅전공 조교수로 재직 중이다. 이화여대 경영학과와 대학원(MIS전공)을 졸업하고, University of Rhode Island에서 경영학박사학위(마케팅전공)를 취득하였다. 한국 딜로이트경영컨설팅에서 컨설턴트로, 미국 Penn State University-Mont Alto에서 조교수로 재직하였다. 주 연구분야는 고객관계관리, 관계마케팅, 디자인경영 등이다.