

제품 네트워크 분석을 이용한 고객의 구매제품 특성 비교 연구*

최일영** · 김재경**†

Product Network Analysis to Analyze the Purchase Behavior of Customers

Il Young Choi** · Jae Kyeong Kim**

■ Abstract ■

As development of information technologies, customer retention has been an important issue in the competing environment. A lot of researches focus on prediction of the churning customers and seeking their characteristics. However, relationships among customers or products have not been considered in existing researches.

In this study, product networks are proposed and analyzed to investigate the differences of network characteristics of products purchased by potential churning customers and those of loyal customers. The product networks are constructed from real product purchase data collected from a Korean department store. We investigated the characteristic differences, such as the degree centrality, degree centralization, and density, of two product networks constructed by potential churning customers and the loyal customers. The results indicate that degree centrality, density and degree centralization of the product network of the loyal customers are higher than those of the potential churning customers. And the promotional products of the department store are resulted to be effective in attracting the loyal customers.

Keyword : Social Network Analysis, Churning Analysis, Product Network, Degree Centrality, Density, Degree Centralization

논문접수일 : 2009년 07월 07일 논문게재확정일 : 2009년 11월 18일

논문수정일(1차 : 2009년 10월 13일)

* 이 논문은 2007년 정부(교육인적자원부)의 재원으로 한국학술진흥재단의 지원을 받아 수행된 연구임(KRF-2007-321-B00051).

** 경희대학교 경영대학/경영연구원

† 교신저자

1. 서론

정보기술의 발전으로 인해 신규 경쟁업체들이 빠르게 시장에 진입하고 있으며, 고객들의 니즈는 복잡하고 다양해지고 있다. 이에 따라 기업들은 치열한 경쟁환경에 노출되어 있으며, 이러한 경영 환경에서 신규고객의 획득보다는 비용이 저렴하고 고객 관계의 강화를 통해 브랜드 충성도를 높이는데 유리한 기존 고객 유지에 대한 중요성이 커지고 있다 [14, 20].

지금까지 기존 고객을 유지하기 위한 연구는 고객의 이탈방지에 초점을 맞추어 진행되어 왔다. 이탈고객이란 일정기간 동안 구매기록이 없는 고객으로 정의되며[5], 많은 연구들은 인구통계학적 데이터와 거래데이터 등을 기반으로 통계분석 및 데이터 마이닝 기법을 이용하여 고객의 이탈에 영향을 주는 요인을 분석하고, 고객의 구매 패턴을 분석하여 고객의 이탈 가능성을 예측하였다. 김상용, 송지연, 이기순[4]은 인구통계학적 정보와 거래 정보를 이용한 로지스틱 회귀분석을 통해 유지고객과 이탈 고객의 변수별 특성차이를 분석하였다. Berson et al.[10]은 의사결정나무(Decision tree) 기법을 이용하여 잠재적 이탈고객을 분석하였으며, Ng and Liu [14]는 데이터베이스와 로그파일을 이용한 다단계 연관 규칙 탐사 기법을 통해 이탈고객을 예측하였다.

최근에는 기존 이탈방지 모형과 병행하여 이탈고객을 관리할 수 있는 사회 네트워크 분석을 이용한 연구가 진행되고 있다. 사회 네트워크 분석은 고객들간의 연결상태 및 연결 구조를 계량적으로 분석할 수 있는 기법으로[1-3], 김재경 등[5]은 고객 네트워크 분석을 통해 충성고객 그룹과 이탈고객 그룹의 관계적 속성을 계량적으로 분석하였다. 그러나 김재경 등[5]의 연구는 충성고객 그룹과 이탈고객 그룹의 연결정도 중심성이 높은 고객이 구매한 제품의 빈도 분석을 통해 구매제품의 차이를 분석하였기 때문에 각 그룹이 구매한 제품간의 연결상태 및 연결 구조의 차이를 반영하지 못하였을 뿐만 아니라 적은 수의 이탈고객을 분석함으로써 다양한 고객들의 특징을 반영하지 못한 한계점이 있다.

따라서 본 연구에서는 매월 구매한 고객을 충성고객으로 정의하고 일정기간 매월 구매하고 그 이후 구매와 비구매를 반복하는 고객을 잠재적 이탈고객으로 정의하여, 충성고객이 구매한 제품 네트워크와 잠재적 이탈고객이 구매한 제품 네트워크의 연결관계 및 연결구조의 차이를 사회 네트워크 분석을 통해 파악하고자 한다. 즉 본 연구의 목적은 고객의 구매 데이터를 활용하여 충성고객이 구매한 제품 네트워크와 잠재적 이탈고객이 구매한 제품 네트워크를 분석하여 충성고객의 구매제품과 잠재적 이탈고객의 구매제품 차이를 파악하는 것으로 국내 L백화점 고객을 대상으로 적용해 보았다.

본 논문의 구성은 다음과 같다. 제 2장에서는 사회네트워크 분석을 살펴보았으며, 제 3장에서는 실제 구매 데이터를 이용하여 충성고객의 구매제품 네트워크와 잠재적 이탈고객의 구매제품 네트워크를 실증적으로 분석하였다. 그리고 제 4장에서는 본 연구의 결론 및 향후 연구방향을 기술하였다.

2. 사회 네트워크 분석

2.1 사회 네트워크 분석

사회 네트워크 분석은 개체간의 관계를 시각적으로 표현하여 개체간 연결상태 및 연결구조의 특성을 계량적으로 분석하는 기법으로[1-3], 교통 네트워크[6], 조직 네트워크[7], 유전 네트워크[9] 등의 구조 분석을 위해 사회과학, 경영학, 응용과학 등 다양한 분야에서 응용되고 있다.

고객의 구매 데이터를 이용하여 제품간 연결상태를 시각적으로 나타내는 방법으로는 장바구니 분석을 통해 도출된 제품들간의 연관 관계를 그래프로 표현하는 방법이 있다. 그러나 장바구니 분석은 연관 관계가 있는 제품이 소수일 경우 제품들의 연결구조를 시각적으로 파악 가능하지만, 연관 관계가 있는 제품이 많을 경우 제품들 간의 구조적 특징을 시각적으로 분석하기에는 적합하지 않다. 따라서 본 연구에서는 제품간 연결상태 및 연결구조를 계량적으로 파악하기 위해 준 연결망(Quasi network)

을 이용하여 제품 네트워크를 구성하였다. 준 연결망을 이용한 제품 네트워크는 구매한 제품들 사이에 직접적인 상호작용의 관계가 없더라도 관계를 인위적으로 설정하여 제품과 제품 사이의 관계를 표현한 네트워크로[1, 5], 노드는 고객들이 구매한 제품을 의미하며 링크는 고객이 구매한 제품간의 관계를 나타낸다. 일반적으로 준 연결망에서 관계를 나타내기 위한 방법으로는 유클리드 거리(Euclidean distance), 동시 발생 빈도(Co-occurrence) 등이 사용된다[1]. 본 연구에서는 제품들간의 관계뿐만 아니라 관계의 시작(Sender)와 끝(Receiver)을 나타내는 제품 네트워크를 구성하기 위하여 고객들이 구매한 제품과 더불어 자주 구매되는 제품을 측정하는 장바구니 분석의 지지도와 신뢰도를 이용하였다. 이렇게 방향을 나타내는 네트워크에서는 관계의 방향에 따라 인디그리(In-degree)와 아웃디그리(Out-degree)로 구분된다[1-3]. 인디그리는 어떤 노드(X)가 다른 노드(Y)에게 화살표를 받는 관계를 의미하고($X \leftarrow Y$), 아웃디그리는 어떤 노드(X)가 다른 노드(Y)에게 화살표를 주는 관계를 의미한다($X \rightarrow Y$)[1].

네트워크 구조를 파악하기 위해 사회 네트워크 분석에서 사용되는 측정 지표로는 중심성(Centrality), 집중도(Centralization) 및 밀도(Density) 등이 있다[1-3, 12]. 중심성은 특정 노드가 전체 네트워크에서 중심에 위치하는 정도를 나타내는 지표로써, 연결정도 중심성(Degree centrality), 근접 중심성(Closeness centrality), 매개 중심성(Betweenness centrality)으로 구분된다. 연결정도 중심성은 한 노드의 지역 중심성을 측정하는 지표이며, 근접 중심성은 한 노드의 전체 중심성을 측정하는 지표이다. 매개 중심성은 네트워크내에서 한 노드와 다른 노드들 사이에서 중재자 역할의 정도를 측정하는 지표이다. 그러나 노드 개수가 다른 네트워크를 절대적 중심성을 이용하여 비교하는 것은 불가능하기 때문에 네트워크간 비교를 위해 상대적 중심성을 측정한다[1-3]. 상대적 연결정도 중심성은 한 노드에 직접적으로 연결되어 있는 노드의 개수를 네트워크 규모로 나눈 값으로 정의되며, 노드 i 의 상대적 연결정도

중심성($C'_D(i)$)은 식 (1)과 같이 계산된다[3].

$$C'_D(i) = d(m_i)/(g-1) \quad (1)$$

여기서 $d(m_i)$ 는 노드 i 와 연결된 노드들의 총수이며, g 는 네트워크에 존재하는 노드들의 총수이다.

상대적 근접 중심성은 한 노드로부터 다른 노드에 도달하기 위해 필요한 최소 단계의 합의 역수에 네트워크 규모를 가중한 값으로 정의되며, 노드 i 의 상대적 근접중심성($C'_C(i)$)은 식 (2)와 같이 계산된다[3].

$$C'_C(i) = (g-1)[\sum_{j=1}^n d_{ij}]^{-1} \quad (2)$$

여기서 d_{ij} 는 두 노드 i 와 j 를 잇는 가장 짧은 경로거리를 의미하며, g 는 네트워크에 존재하는 노드들의 총수이다.

상대적 매개 중심성은 네트워크내에서 한 쌍의 노드들 사이의 존재하는 최단거리 경로 중에서 특정 노드를 통과해야 하는 경로의 숫자 비율을 최대 가능한 매개 중심성 값으로 나누어서 얻은 값으로 정의되며, 노드 i 의 상대적 매개 중심성($C'_B(i)$)은 식 (3)과 같이 계산된다[3].

$$C'_B(i) = (\sum_{j < k} g_{jk}(i)/g_{jk})/[(g-1)(g-2)/2] \quad (3)$$

여기서 g_{jk} 는 네트워크내 특정 두 노드(j 와 k) 사이에 존재하는 최단거리경로들의 경우의 숫자이고, 는 두 노드 j 와 $k(j \neq k)$ 사이에 존재하는 노드 i 를 경유하는 횟수를 말한다. 또한 g 는 네트워크에 존재하는 노드들의 총수이다.

집중도는 네트워크 전체가 한 노드로 집중되는 정도를 표현하는 지표로써, 연결정도 집중도(Degree centralization), 근접 집중도(Closeness centralization), 매개 집중도(Betweenness centralization)로 구분된다[1-3]. 연결정도 집중도는 각 노드간의 연결 정도에 의존하여 전체 네트워크의 집중화 정도를 측정하는 것으로 식 (4)와 같이 계산된다[3].

$$C_D = (\sum_{i=1}^g [C_D(n^*) - C_D(n_i)]) / [(g-1)(g-2)] \quad (4)$$

여기서 $C_D(n^*)$ 는 네트워크의 가장 높은 연결정도 중심성 값이며, $C_D(n_i)$ 는 노드 i 의 연결정도 중심성 값이다. g 는 네트워크에 존재하는 노드들의 총수이다.

근접 집중도는 각 노드의 거리에 기반하여 전체 네트워크의 집중화 정도를 측정하는 지표로 식 (5)와 같이 계산된다[3].

$$C_C = (\sum_{i=1}^g [C_C(n^*) - C_C(n_i)]) / [(g-1)(g-2)/(2g-3)] \quad (5)$$

여기서 $C_C(n^*)$ 는 네트워크의 가장 높은 근접 중심성 값이며, $C_C(n_i)$ 는 노드 i 의 근접 중심성 값이다. g 는 네트워크에 존재하는 노드들의 총수이다[3].

매개 집중도는 각 노드의 매개성을 기반하여 전체 네트워크의 집중화 정도를 측정하는 지표로 식 (6)과 같이 계산된다[3].

$$C_B = (\sum_{i=1}^g [C_B(n^*) - C_B(n_i)]) / [(g-1)^2(g-2)] \quad (6)$$

여기서 $C_B(n^*)$ 는 네트워크의 가장 높은 매개 중심성 값이며, $C_B(n_i)$ 는 노드 i 의 매개 중심성 값이다. g 는 네트워크에 존재하는 노드들의 총수이다.

밀도는 네트워크 노드들 사이의 연결된 정도로, 밀도가 높은 네트워크는 정보의 교류가 활성화되어 정보의 확산이 빠르다. 일반적으로 밀도(G)는 식 (7)과 같다[3].

$$G = k / [g(g-1)] \quad (7)$$

여기서, k 는 네트워크내에 존재하는 링크의 총수이고, g 는 네트워크에 존재하는 노드들의 총수이다.

본 연구에서는 충성고객이 구매한 제품 네트워크와 잠재적 이탈고객이 구매한 제품 네트워크의 구조적 차이를 분석하기 위하여, 위에서 제시한 여러 개의 측정지표 중 상대적 연결정도 중심성, 연결정도 집중도 및 밀도를 사용하였다. 제품 네트워크에

서 연결정도 중심성은 고객들의 제품 선호도를 나타낸다. 특히 아웃도그리 연결정도 중심성이 높은 제품은 집객효과가 큰 제품이며, 인디그리 연결정도 중심성이 높은 제품은 고객에게 인기있는 제품을 의미한다. 따라서 충성고객과 잠재적 이탈고객이 선호하는 제품의 차이를 분석하기 위하여 상대적 연결정도 중심성을 사용하였다. 또한 제품 네트워크에서 연결정도 집중도는 고객들이 구매한 제품 중 특정 제품을 중심으로 얼마나 결속되어 있는가를 나타내는 지표로, 제품 네트워크에서 연결정도 중심성이 높을수록 특정 제품에 대한 고객의 선호가 크다는 것을 의미한다. 밀도는 고객들이 구매한 다양한 제품 중 실제로 선호하는 제품들의 비율로, 제품 네트워크에서 밀도가 높을수록 고객의 선호 제품이 다양하다는 것을 의미한다. 따라서 충성고객과 잠재적 이탈고객의 제품 네트워크에서 연결정도 집중도와 밀도를 살펴봄으로써 충성고객과 잠재적 이탈고객의 제품 선호에 대한 차이를 측정할 수 있다. 반면에 제품 네트워크에서 근접중심성은 고객이 구매한 제품이 어느 정도 네트워크의 중심에 있는가를 나타내는 개념으로 충성고객과 잠재적 이탈고객의 제품 네트워크의 구조적 차이를 분석하는데 적합하지 않다고 판단된다. 매개중심성은 고객의 선호하는 제품을 다른 제품으로 구매를 유인할 수 있는가를 나타내는 개념이다. 예를 들어 제품 네트워크에서 제품 X 와 제품 Y 가 제품 Z 를 통해 연결되어 있다면, 제품 X 를 선호하는 고객에게 제품 Y 의 구매를 유도하기 위해서는 제품 Y 를 직접적으로 권유하기 보다는 제품 Z 를 권유함으로써 간접적으로 제품 Y 의 구매를 유도할 수 있다. 그러나 매개중심성 또한 충성고객과 잠재적 이탈고객이 구매한 제품 네트워크의 구조적 차이를 분석하는데 적합하지 않다고 판단되므로 본 연구에서는 사용하지 않기로 한다.

2.2 제품 네트워크 분석

2.2.1 제품 네트워크 분석 프로세스

본 연구에서는 고객의 구매 데이터를 이용하여

충성고객의 제품 네트워크와 잠재적 이탈고객의 제품 네트워크를 구성하고, 네트워크 분석을 통해 충성고객과 잠재적 이탈고객의 제품 네트워크 사이의 구조적 차이를 분석한다. 즉 충성고객의 제품 네트워크와 잠재적 이탈고객의 제품 네트워크의 연결상태, 연결구조 등을 분석함으로써, 충성고객의 제품 네트워크와 잠재적 이탈고객의 제품 네트워크의 차별적 특성을 구하고자 한다.

본 연구가 제안하는 제품 네트워크는 고객이 선호하는 제품들간의 관계를 나타내는 네트워크로, 제품 네트워크 분석 프로세스는 김재경 등[5]이 제안한 고객 네트워크 분석 프로세스를 기반으로 하여 제품 네트워크를 분석하는데 적합하도록 수정하였다. 제품 네트워크 프로세스는 고객 선호 프로파일 생성, 제품 네트워크 구성 및 네트워크 구조 분석 등 3단계 구성된다. 고객 선호 프로파일 생성단계에서는 충성고객과 잠재적 이탈고객의 구매 데이터를 이용하여 고객-제품 매트릭스 형태로 충성 고객과 잠재적 이탈고객의 제품 선호도를 표현한다. 네트워크 구성단계에서는 장바구니 분석(Market Basket Analysis)의 지지도(Support)와 신뢰도(Confidence)를 이용하여 특정 임계치(Threshold) 이상인 제품들을 링크로 연결한다. 마지막 단계에서는 제품 네트

워크의 중심성, 집중도, 밀도 계산을 통해 충성고객이 구매한 제품과 잠재적 이탈고객이 구매한 제품의 특징을 파악한다.

2.2.2 고객 선호 프로파일 생성

고객 선호 프로파일(Customer preference profile)은 인구통계학적 데이터, 메뉴, 콘텐츠, 제품 이용, 구매 정보 및 마케팅 피드백 정보 등 다양한 고객 정보를 바탕으로 한 고객 선호를 구조화한 표현이다. 본 연구에서는 Sarwar et al.[16-18]의 연구처럼 고객의 구매 데이터를 이용하여 i 번째 고객이 특정 기간에 j 번째 제품을 구매 하였으면 1, 그렇지 않으면 0의 값을 입력한 고객-제품 매트릭스 형태로 고객 선호 프로파일을 생성하였다. <표 1>은 충성고객과 이탈고객의 고객-제품 매트릭스 예이다.

2.2.3 제품 네트워크 구성

제품 네트워크는 고객이 선호하는 제품 사이의 관계가 있고 없음을 나타낸 네트워크이다. 예를 들어 고객이 제품 X 와 제품 Y 를 구매하면, 두 제품은 서로 관계가 있다고 가정하여 제품 X 와 제품 Y 를 연결함으로써 네트워크를 구성할 수 있다. 그러나 구매 제품들이 관계가 있다고 가정하면, 구매 제

<표 1> 고객-제품 매트릭스

(a) 충성고객

	기초 화장품	디자이너부틱	란제리	명품잡화	시티웨어	니트	뷰티상품	스포츠화
L_{01}	0	1	1	0	0	1	1	0
L_{02}	1	0	0	0	0	1	1	0
L_{03}	0	0	1	1	0	1	0	1
L_{04}	0	1	0	1	1	0	1	0
L_{05}	1	1	1	0	0	0	0	1

(b) 잠재적 이탈고객

	기초 화장품	디자이너부틱	란제리	명품잡화	시티웨어	니트	뷰티상품	스포츠화
D_{01}	0	1	1	0	0	1	0	1
D_{02}	1	1	0	1	1	1	1	0
D_{03}	1	1	0	0	1	1	0	0
D_{04}	1	1	0	0	0	0	0	1
D_{05}	1	0	1	1	0	1	0	1

품 모두가 링크로 연결되는 문제점이 있기 때문에 본 연구에서는 제품 X와 제품 Y를 구매한 고객이 일정 비율 이상이고, 제품 X를 구매한 고객들 중 제품 Y가 구매한 고객이 일정 비율 이상인 제품들만 관계가 있다고 가정하여 제품들을 링크로 연결하였다.

구매된 제품들 사이의 선호도를 측정하는 방법은 지지도와 신뢰도가 있다. 지지도는 전체 고객 중 제품 X와 제품 Y를 구매한 고객의 비율을 측정하는 지표로써, 식 (8)과 같이 계산된다[8, 19].

$$S = P(X \cap Y) = \frac{\text{제품 X와 제품 Y를 구매한 고객수}}{\text{전체 고객수}(N)} \quad (8)$$

신뢰도는 제품 X를 구매한 고객 중 제품 Y를 구매한 고객을 지표로, 식 (9)와 같이 계산된다[8, 19].

$$C = P(Y|X) = \frac{P(X \cap Y)}{P(X)} \quad (9)$$

= 제품 X와 제품 Y를 구매한 고객수 /
제품 X를 구매한 고객 수

지지도는 제품들간의 관계 유무에 대한 정보를 제공하며, 신뢰도는 제품들간의 관계 방향을 알려준다. 따라서 본 연구에서는 제품들 사이의 관계뿐 아니라 관계의 방향을 나타내는 제품 네트워크를 구성하기 위하여 지지도와 신뢰도를 동시에 고려하였으며, 제품 네트워크는 다음과 같이 두 단계의 과정을 거쳐 구성된다. 단계 1은 제품들 사이의 지지도 및 신뢰도를 계산한다. <표 2>는 <표 1>의 제품에 대한 고객 선호로부터 제품들간의 지지도 및 신뢰도를 계산한 것이다. 예를 들어 란제리와 스포츠화에 대한 지지도와 신뢰도를 살펴보면 란제리를 구매한 충성고객은 L_{01} , L_{03} , L_{05} 이며, 스포츠화를 구

<표 2> 구매제품 지지도 및 신뢰도

(a) 충성고객

전항 값	후항 값	지지도	신뢰도	전항 값	후항 값	지지도	신뢰도
스포츠화	란제리	40.0%	100.0%	니트	기초화장품	20.0%	33.3%
니트	뷰티상품	40.0%	66.7%	니트	명품잡화	20.0%	33.3%
니트	란제리	40.0%	66.7%	니트	스포츠화	20.0%	33.3%
디자이너부틱	뷰티상품	40.0%	66.7%	니트	디자이너부틱	20.0%	33.3%
디자이너부틱	란제리	40.0%	66.7%	디자이너부틱	기초화장품	20.0%	33.3%
란제리	스포츠화	40.0%	66.7%	디자이너부틱	명품잡화	20.0%	33.3%
란제리	니트	40.0%	66.7%	디자이너부틱	스포츠화	20.0%	33.3%
란제리	디자이너부틱	40.0%	66.7%	디자이너부틱	니트	20.0%	33.3%
뷰티상품	니트	40.0%	66.7%	란제리	기초화장품	20.0%	33.3%
뷰티상품	디자이너부틱	40.0%	66.7%	란제리	명품잡화	20.0%	33.3%
시티웨어	명품잡화	20.0%	100.0%	란제리	뷰티상품	20.0%	33.3%
시티웨어	뷰티상품	20.0%	100.0%	뷰티상품	시티웨어	20.0%	33.3%
시티웨어	디자이너부틱	20.0%	100.0%	뷰티상품	기초화장품	20.0%	33.3%
기초화장품	스포츠화	20.0%	50.0%	뷰티상품	명품잡화	20.0%	33.3%
기초화장품	니트	20.0%	50.0%	뷰티상품	란제리	20.0%	33.3%
기초화장품	뷰티상품	20.0%	50.0%	기초화장품	시티웨어	0.0%	0.0%
기초화장품	디자이너부틱	20.0%	50.0%	기초화장품	명품잡화	0.0%	0.0%
기초화장품	란제리	20.0%	50.0%	니트	시티웨어	0.0%	0.0%
명품잡화	시티웨어	20.0%	50.0%	란제리	시티웨어	0.0%	0.0%
명품잡화	스포츠화	20.0%	50.0%	명품잡화	기초화장품	0.0%	0.0%
명품잡화	니트	20.0%	50.0%	뷰티상품	스포츠화	0.0%	0.0%
명품잡화	뷰티상품	20.0%	50.0%	스포츠화	시티웨어	0.0%	0.0%
명품잡화	디자이너부틱	20.0%	50.0%	스포츠화	뷰티상품	0.0%	0.0%
명품잡화	란제리	20.0%	50.0%	시티웨어	기초화장품	0.0%	0.0%
스포츠화	기초화장품	20.0%	50.0%	시티웨어	스포츠화	0.0%	0.0%
스포츠화	명품잡화	20.0%	50.0%	시티웨어	니트	0.0%	0.0%
스포츠화	니트	20.0%	50.0%	시티웨어	란제리	0.0%	0.0%
스포츠화	디자이너부틱	20.0%	50.0%				

(b) 잠재적 이탈고객

전항 값	후항 값	지지도	신뢰도	전항 값	후항 값	지지도	신뢰도
기초화장품	디자이너부틱	60.0%	75.0%	니트	뷰티상품	20.0%	25.0%
기초화장품	니트	60.0%	75.0%	디자이너부틱	뷰티상품	20.0%	25.0%
니트	디자이너부틱	60.0%	75.0%	디자이너부틱	란제리	20.0%	25.0%
니트	기초화장품	60.0%	75.0%	디자이너부틱	명품잡화	20.0%	25.0%
디자이너부틱	기초화장품	60.0%	75.0%	스포츠화	명품잡화	20.0%	33.3%
디자이너부틱	니트	60.0%	75.0%	란제리	명품잡화	20.0%	50.0%
기초화장품	시티웨어	40.0%	50.0%	란제리	디자이너부틱	20.0%	50.0%
기초화장품	명품잡화	40.0%	50.0%	란제리	기초화장품	20.0%	50.0%
기초화장품	스포츠화	40.0%	50.0%	명품잡화	뷰티상품	20.0%	50.0%
니트	란제리	40.0%	50.0%	명품잡화	란제리	20.0%	50.0%
니트	시티웨어	40.0%	50.0%	명품잡화	시티웨어	20.0%	50.0%
니트	명품잡화	40.0%	50.0%	명품잡화	스포츠화	20.0%	50.0%
니트	스포츠화	40.0%	50.0%	명품잡화	디자이너부틱	20.0%	50.0%
디자이너부틱	시티웨어	40.0%	50.0%	시티웨어	뷰티상품	20.0%	50.0%
디자이너부틱	스포츠화	40.0%	50.0%	시티웨어	명품잡화	20.0%	50.0%
스포츠화	란제리	40.0%	66.7%	뷰티상품	시티웨어	20.0%	100.0%
스포츠화	디자이너부틱	40.0%	66.7%	뷰티상품	명품잡화	20.0%	100.0%
스포츠화	기초화장품	40.0%	66.7%	뷰티상품	디자이너부틱	20.0%	100.0%
스포츠화	니트	40.0%	66.7%	뷰티상품	기초화장품	20.0%	100.0%
란제리	스포츠화	40.0%	100.0%	뷰티상품	니트	20.0%	100.0%
란제리	니트	40.0%	100.0%	란제리	뷰티상품	0.0%	0.0%
명품잡화	기초화장품	40.0%	100.0%	란제리	시티웨어	0.0%	0.0%
명품잡화	니트	40.0%	100.0%	뷰티상품	란제리	0.0%	0.0%
시티웨어	디자이너부틱	40.0%	100.0%	뷰티상품	스포츠화	0.0%	0.0%
시티웨어	기초화장품	40.0%	100.0%	스포츠화	뷰티상품	0.0%	0.0%
시티웨어	니트	40.0%	100.0%	스포츠화	시티웨어	0.0%	0.0%
기초화장품	뷰티상품	20.0%	25.0%	시티웨어	란제리	0.0%	0.0%
기초화장품	란제리	20.0%	25.0%	시티웨어	스포츠화	0.0%	0.0%

매한 충성고객은 L_{03} , L_{05} 이다. 따라서 란제리 \Rightarrow 스포츠화[지지도 = 40%, 신뢰도 = 66.7%]이며, 스포츠화 \Rightarrow 란제리[지지도 = 40%, 신뢰도 = 100%]이다.

단계 2는 제품 사이의 지지도와 유사도가 특정 임계치(α , ρ)이상인 값을 식 (10)과 같이 1로 정의 하여 제품들을 링크로 연결한다.

$$R_{x,y} = \begin{cases} 1, & \text{if } Support(x, y) \geq \sigma \text{ and } Confidence(x, y) \geq \rho \\ 0, & \text{otherwise} \end{cases} \quad (10)$$

여기서 $R_{x,y}$ 는 제품 x 와 제품 y 의 관계를 나타 내고, $Support(x, y)$ 와 $Confidence(x, y)$ 는 제품 x 와 제품 y 의 지지도와 신뢰도를 의미한다.

<표 3>은 <표 2>의 제품 사이의 지지도와 신뢰도에 대한 임계치를 20%와 50%로 설정하여 그 이

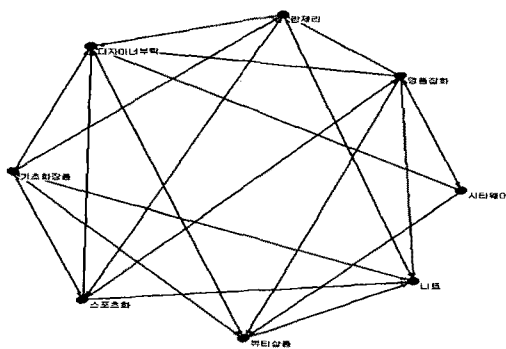
상의 값을 가진 제품들간의 인접 매트릭스를 표현한 것이다. 예를 들어 충성고객이 구매한 란제리, 니트, 스포츠화에 대한 지지도와 신뢰도를 살펴보면 란제리 \Rightarrow 니트[지지도 = 40%, 신뢰도 = 66.7%], 란제리 \Rightarrow 스포츠화[지지도 = 40%, 신뢰도 = 66.7%], 니트 \Rightarrow 스포츠화[지지도 = 20%, 신뢰도 = 33.3%], 니트 \Rightarrow 란제리[지지도 = 40%, 신뢰도 = 66.7%], 스포츠화 \Rightarrow 니트[지지도 = 20%, 신뢰도 = 50%]이다. 란제리 \Rightarrow 니트, 란제리 \Rightarrow 스포츠화, 니트 \Rightarrow 란제리, 스포츠화 \Rightarrow 란제리, 스포츠화 \Rightarrow 니트는 지지도와 신뢰도에 대한 임계치 이상이므로 $R_{란제리,니트} = 1$, $R_{란제리,스포츠화} = 1$, $R_{니트,란제리} = 1$, $R_{스포츠화,란제리} = 1$, $R_{스포츠화,니트} = 1$ 이다. 반면에 니트 \Rightarrow 스포츠화는 지지도가 20% 이상이지만 신뢰도가 50% 미만이기 때문에 $R_{니트,스포츠화} = 0$ 이다.

<표 3> 제품 인접 매트릭스(지지도 20%, 신뢰도 50%)
(a) 충성고객

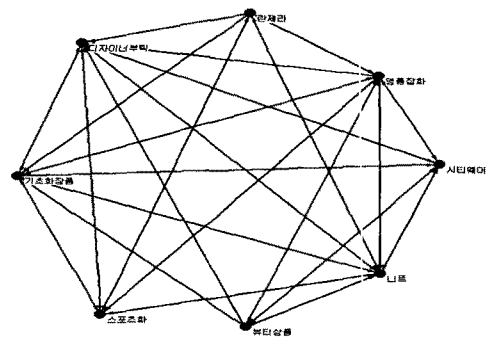
	기초화장품	디자이너부틱	란제리	명품잡화	시티웨어	니트	뷰티상품	스포츠화
기초화장품	0	1	1	0	0	1	1	1
디자이너부틱	0	0	1	0	0	0	1	0
란제리	0	1	0	0	0	1	0	1
명품잡화	0	1	1	0	1	1	1	1
시티웨어	0	1	0	1	0	0	1	0
니트	0	0	1	0	0	0	1	0
뷰티상품	0	1	0	0	0	1	0	0
스포츠화	1	1	1	1	0	1	0	0

(b) 잠재적 이탈고객

	기초화장품	디자이너부틱	란제리	명품잡화	시티웨어	니트	뷰티상품	스포츠화
기초화장품	0	1	0	1	1	1	0	1
디자이너부틱	1	0	0	0	1	1	0	1
란제리	1	1	0	1	0	1	0	1
명품잡화	1	1	1	0	1	1	1	1
시티웨어	1	1	0	1	0	1	1	0
니트	1	1	1	1	1	0	0	1
뷰티상품	1	1	0	1	1	1	0	0
스포츠화	1	1	1	0	0	1	0	0



(a) 충성고객



(b) 잠재적 이탈고객

[그림 1] 제품 네트워크

[그림 1]은 <표 3>을 이용하여 제품 네트워크를 그래프로 표현한 예로 충성고객과 잠재적 이탈고객이 구매한 제품은 노드로 표현하였으며, 지지도와 신뢰도의 임계치를 각각 20%와 50%로 설정하여 그

이상인 제품을 링크로 연결하였다. 예를 들어 충성고객이 구매한 란제리, 니트, 스포츠화에 대해 $R_{란제리,니트} = 1$, $R_{란제리,스포츠화} = 1$, $R_{니트,란제리} = 1$, $R_{스포츠화,란제리} = 1$, $R_{스포츠화,니트} = 1$ 이므로 링크로 연결되

었으나, $R_{\text{네트워크, 스포츠화}} = 0$ 이므로 링크로 연결되지 않았다.

2.2.4 네트워크 구조 분석

네트워크 구조 분석은 구성된 충성고객의 제품 네트워크와 잠재적 이탈고객의 제품 네트워크에 대한 밀도, 연결정도 집중도 분석 및 연결정도 중심성 분석을 통해 각 네트워크의 구조적 특징을 발견하는 단계이다. 여기서 밀도와 연결정도 집중도는 네트워크에 대한 지표로써 밀도가 높다는 것은 다양한 제품을 선호한다는 것을 의미한다. 연결정도 집중도가 높다는 것은 소수의 제품에 대해 고객의 선호가 높다는 것을 의미한다. 또한 연결정도 중심성은 노드에 대한 지표로써 특정 제품의 연결정도 중심성이 높다는 것은 고객들은 특정 한 제품과 다수의 다른 제품을 구매한다는 것을 의미하는 것으로 아웃디그리 연결정도 중심성은 제품의 집객효과를 나타내고 인디그리 연결정도는 제품의 인기도를 의미한다. 따라서 충성고객이 구매한 제품 네트워크와 잠재적 이탈고객이 구매한 제품 네트워크의 구조 분석을 통해 차이가 발생하면 제품 분석을 통해 잠재적 이탈고객을 충성고객으로 유도할 수 있는 마케팅 전략이 가능하게 된다.

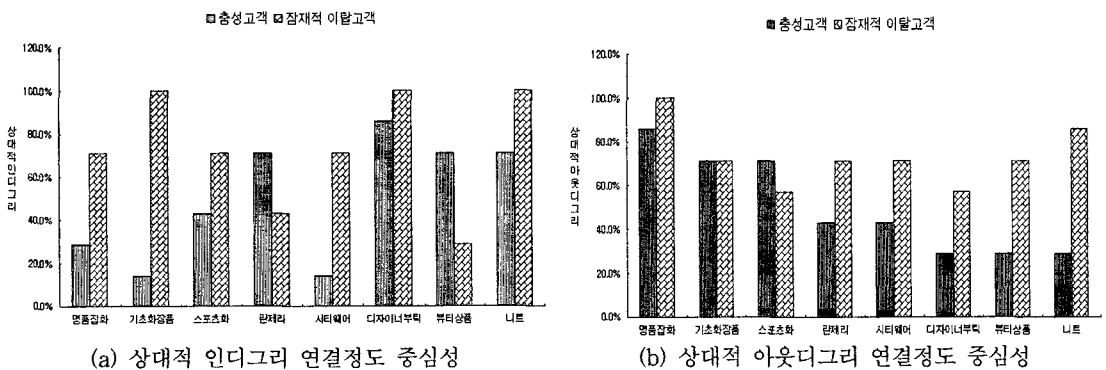
<표 4>는 [그림 1]을 이용하여 충성고객의 제품 네트워크와 잠재적 이탈고객의 제품 네트워크에 대한 밀도, 연결정도 집중도를 나타낸 예로 잠재적 이

탈고객의 제품 네트워크가 충성고객의 제품 네트워크보다 밀도는 높은 반면에 연결정도 집중도는 낮다. 이는 잠재적 이탈고객은 다수의 제품을 구매하지만, 특정 제품에 대한 선호가 충성고객보다 상대적으로 명확하지 않다는 것을 의미한다.

<표 4> 제품 네트워크 밀도 연결정도 중심성, 연결정도 집중도

구분	충성고객의 제품 네트워크	잠재적 이탈고객
노드 수	8	8
링크 수	28	41
밀도	50.00%	73.21%
아웃디그리 연결정도 집중도	40.82%	30.61%
인디그리 연결정도 집중도	40.82%	30.61%

충성고객과 잠재적 이탈고객이 선호하는 제품을 살펴보기 위해 연결정도 중심성을 살펴보면 [그림 2]와 같으며, 디자이너부탁은 충성고객의 제품 네트워크와 잠재적 이탈고객의 제품 네트워크에서 인디그리 연결정도 중심성이 높다. 또한 명품잡화는 충성고객의 제품 네트워크와 잠재적 이탈고객의 제품 네트워크에서 아웃디그리 연결정도 중심성이 높다. 그러나 뷰티상품 및 란제리의 인디그리 연결정도 중심성과 스포츠화의 아웃디그리 연결정도 중심성은 충성고객의 제품 네트워크에서 잠재적 이탈고객



[그림 2] 상대적 연결정도 중심성

의 제품 네트워크보다 높다. 즉 스포츠화는 충성고객들의 구매를 촉진하는 미끼상품의 역할을 하며, 뷰티상품 및 라제리는 충성고객에게 인기가 높은 제품이라 해석할 수 있다. 따라서 이러한 충성고객의 특성을 이용하여 카탈로그 첫 장에 스포츠화 배치 등의 마케팅 전략을 실행한다면 고객의 방문 유도를 통해 매출을 증대시킬 수 있다고 판단된다.

3. 충성고객의 제품 네트워크/잠재적 이탈고객의 제품 네트워크 실증 분석

3.1 조사방법

본 연구의 실험을 위해 사용된 데이터는 2007년 1월부터 2007년 12월까지 1년동안 국내 L백화점의 고객 구매 데이터이다. 구매 데이터는 본점을 포함한 24개 지점의 VIP 고객 1,000명, 최우수 고객 1,000명, 우수 고객 1,000명, 일반 고객 999명, 일회성 고객 996명, 휴먼 고객 993명 등 총 5,988명의 고객과 식품을 포함한 331개의 제품군으로 구성되어 있으며, 데이터 양은 400,935개의 레코드이다. 그러나 백화점 식품관의 경우 인근 거주고객이 주로 이용하므로, 충성고객의 제품 네트워크와 잠재적 이탈고객의 제품 네트워크의 차별적 특성을 분석하는데 있어 고객의 식품관 구매 내역이 포함된 데이터를 이용하는 것은 적합하지 않다고 판단하였다. 따라서 본 연구에서는 251,502개의 식품관 구매 데이터를 제외하고 246개 제품군에 대한 5,580명 고객의 구매 데이터인 149,433개의 레코드를 대상으로 분석하였다. 또한 충성고객과 잠재적 이탈고객을 선정하기 위하여 충성고객은 2007년 1월부터 12월까지 매월 구매한 고객으로 정의하였으며, 잠재적 이탈고객은 1월에서 6월까지 매월 구매하고 7월 이후부터는 매월 구매하지는 않았지만, 몇 개월마다 구매가 발생한 고객으로 정의하였다.

본 연구에서는 제품 네트워크를 구성하기 위하여 2007년 1월부터 6월까지 구매한 고객을 대상으로 충

성고객과 잠재적 이탈고객을 무작위로 각각 250명을 선정하였으며, 선정된 충성고객과 잠재적 이탈고객의 구매 데이터를 Excel 파일로 변환하였다. 변환된 파일을 Excel의 피벗테이블 기능을 이용하여 충성고객과 잠재적 이탈고객의 선호 프로파일을 생성하였으며, 데이터 마이닝 툴인 Clementine 10.0을 사용하여 충성고객과 잠재적 이탈고객의 선호 프로파일로부터 제품 사이의 지지도와 신뢰도를 계산하였다. 또한 계산된 지지도와 신뢰도를 사회 네트워크 분석 프로그램인 Ucinet 6.0을 이용하여 충성고객의 제품 네트워크와 잠재적 이탈고객의 제품 네트워크를 구성하여 분석하였다.

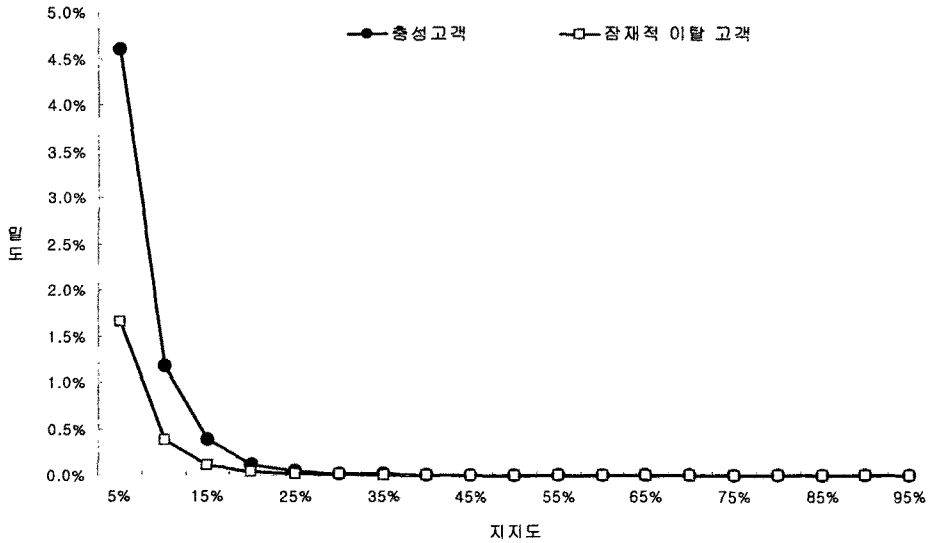
3.2 충성고객의 구매제품 네트워크/잠재적 이탈고객의 구매제품 네트워크 분석

구매 데이터를 이용하여 제품 네트워크를 구성할 때 제품간의 관계를 나타내기 위한 최적의 임계치는 데이터 특성에 따라 달라진다. 본 연구에서는 충성고객의 제품 네트워크와 잠재적 이탈고객의 제품 네트워크를 비교하였을 때 밀도의 차이가 가장 크게 나타나는 임계치를 기준으로 분석하였다.

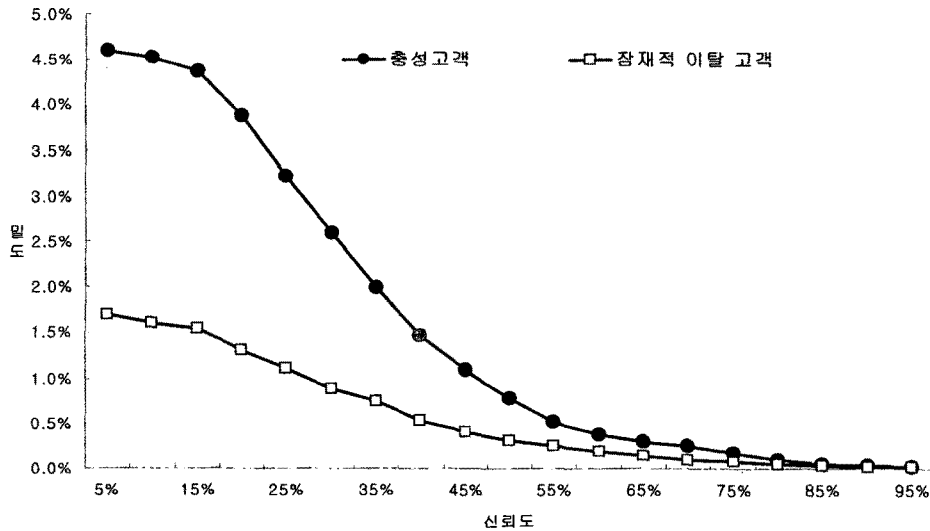
[그림 3]은 2007년 1월부터 6월까지 6개월 동안 구매 데이터를 이용하여 구성한 제품 네트워크에서 지지도별로 충성고객의 제품 네트워크와 잠재적 이탈고객의 제품 네트워크의 밀도 차이를 나타낸 것으로 모든 지지도에서 충성고객의 제품 네트워크가 잠재적 이탈고객의 제품 네트워크의 밀도보다 크거나 같았으며, 특히 지지도 5%에서 밀도의 차이가 가장 크게 나타났다.

[그림 4]는 지지도 5%에서 신뢰도별로 충성고객의 구매제품 네트워크와 잠재적 이탈고객의 제품 네트워크의 밀도 차이를 나타낸 것으로 신뢰도 10%에서 밀도의 차이가 가장 크게 나타났다.

<표 5>는 지지도 5%, 신뢰도 10% 수준에서 충성고객의 제품 네트워크와 잠재적 이탈고객의 제품 네트워크의 평균 연결정도 중심성의 차이가 있는지 독립표본 t-검정한 결과로 유의확률은 0.000으로 나타났다. 따라서 유의 수준이 0.05보다 작으므로 충



[그림 3] 지지도별 밀도

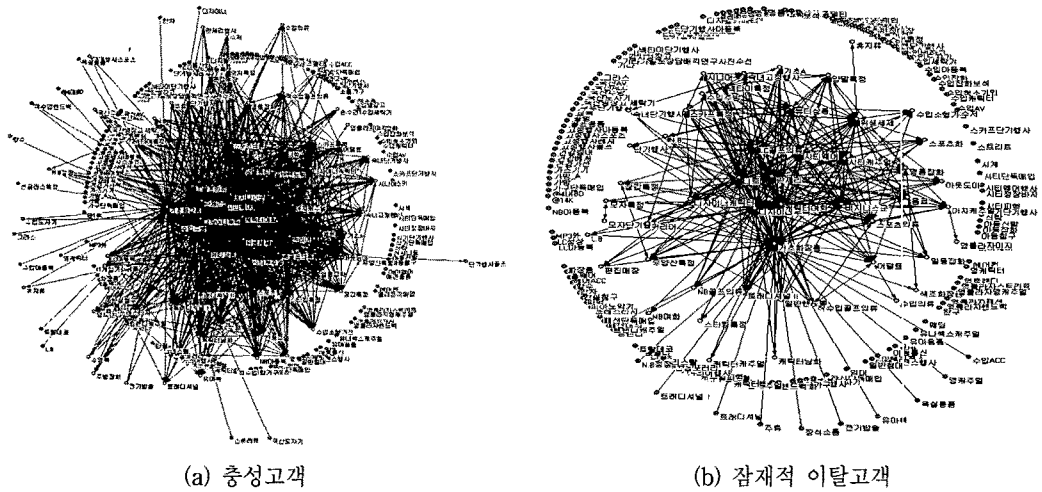


[그림 4] 지지도 5%에서 신뢰도별 밀도

<표 5> 고객의 구매제품 네트워크의 독립표본 t-검정 결과

측정치	그룹	평균	t값	유의확률
상대적 아웃디그리 연결정도 중심성	충성고객	4.543	4.594*	0.000*
	잠재적 이탈고객	1.614		
상대적 인디그리 연결정도 중심성	충성고객	4.543	4.430*	0.000*
	잠재적 이탈고객	1.614		

주) * p < 0.05.



[그림 5] 제품 네트워크

성고객의 구매제품 네트워크와 잠재적 이탈고객의 구매제품 네트워크의 연결정도 중심성의 차이가 통계학적으로 유의한 것으로 나타났다. 따라서 본 연구는 지지도 5%, 신뢰도 10% 수준에서 충성고객의 구매제품 네트워크와 잠재적 이탈고객의 구매제품 네트워크를 분석하였다.

[그림 5]는 지지도 5%, 신뢰도 10% 수준에서 구성된 충성고객의 구매제품 네트워크와 잠재적 이탈고객의 구매제품 네트워크로 충성고객의 구매제품 네트워크가 잠재적 이탈고객의 구매제품 네트워크보다 더 복잡한 네트워크 형태를 나타낸다.

<표 6> 제품 네트워크 특징

구분	충성고객	잠재적 이탈고객
노드 수	199	196
링크 수	1,790	617
고립노드 수	113	135
밀도	4.50%	1.60%
아웃디그리 연결정도 집중도	28.40%	21.10%
인디그리 연결정도 집중도	37.60%	27.80%

지지도 5%, 신뢰도 10% 수준에서 충성고객의 제품 네트워크와 잠재적 이탈고객의 제품 네트워크의

구조적 특징을 살펴보면 <표 6>과 같이, 충성고객의 제품 네트워크가 잠재적 이탈고객의 제품 네트워크보다 고립노드의 수가 적을 뿐만 아니라 집중도가 높음을 알 수 있다. 이는 충성고객들은 선호하는 제품이 다양하며, 그 중 특정 제품에 대한 선호가 크다고 판단할 수 있다.

그러나 충성고객의 제품 네트워크와 잠재적 이탈고객의 제품 네트워크간의 연결상태 및 연결구조가 의미 있는 차이를 가지기 위해서는 일정 값 이상의 지지도와 신뢰도를 임계치로 설정하여야 한다. 일반적으로 지지도와 신뢰도가 높으면 제품 네트워크의 고립노드의 수는 증가하여 제품들의 연결정도 중심성은 낮아진다. 반대로 지지도와 신뢰도가 낮으면 고립노드의 수는 감소하여 제품들의 연결정도 중심성은 높아진다. 따라서 임계치가 높아질수록 연결정도 중심성이 낮은 제품은 고립노드로 전환되므로 연결정도 중심성이 높은 제품을 중심으로 충성고객의 제품 네트워크와 잠재적 이탈고객의 제품 네트워크를 비교하면 두 네트워크간의 의미 있는 구조적 차이를 분석할 수 있다.

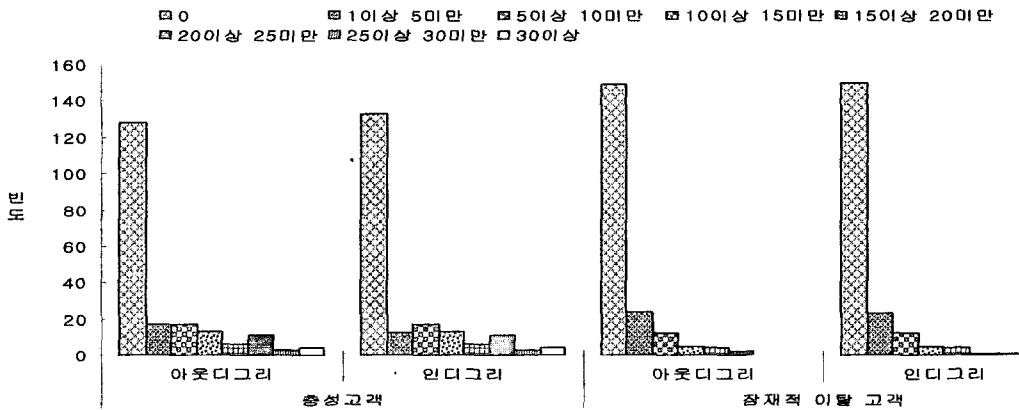
충성고객의 제품 네트워크와 잠재적 이탈고객의 제품 네트워크에서 연결정도 중심성 분포를 살펴보면 [그림 6]과 같이 상대적 연결정도 중심성이 15 이상인 제품의 비중은 충성고객의 구매제품 네트워

크에서 약 12%를 차지하고 있는 반면에 잠재적 이탈고객의 구매제품 네트워크에서는 약 3%를 차지하고 있다.

각 제품 네트워크에서 연결정도 중심성이 상위

10위인 제품을 살펴보면, <표 7>과 같이 아웃디그리 연결정도 중심성이 높은 제품이 인디그리 연결정도 중심성 역시 높게 나왔다.

이러한 구조적 차이를 가져오는 요인은 집객효과



[그림 6] 상대적 연결정도 중심성 분포

<표 7> 상위 10위 연결정도 중심성

(a) 충성고객

순위	구매제품	상대적 아웃디그리 연결정도 중심성	구매제품	상대적 인디그리 연결정도 중심성
1	란제리	32.828	란제리	32.828
2	기초화장품	31.818	기초화장품	41.919
3	디자인너부틱	31.313	디자인너부틱	33.333
4	비즈니스군	30.303	비즈니스군	30.303
5	시티웨어	29.798	시티웨어	29.798
6	캐릭터여화	29.293	캐릭터여화	29.293
7	위생세제	27.778	위생세제	27.778
8	뷰티상품	24.242	뷰티상품	24.242
9	양말특정	24.242	양말특정	24.242
10	LC골프의류	23.737	LC골프의류	23.737

(b) 잠재적 이탈고객

순위	구매제품	상대적 아웃디그리 연결정도 중심성	구매제품	상대적 인디그리 연결정도 중심성
1	기초화장품	22.564	기초화장품	29.231
2	디자인너부틱	21.538	디자인너부틱	21.538
3	란제리	17.949	란제리	17.949
4	캐릭터여화	17.436	캐릭터여화	17.436
5	시티웨어	15.897	시티웨어	15.897
6	니트	15.897	니트	15.897
7	LC골프의류	12.821	LC골프의류	12.821
8	위생세제	12.308	위생세제	12.308
9	시티캐주얼	11.795	시티캐주얼	11.795
10	뷰티상품	10.769	뷰티상품	10.769

가 큰 제품이 인기가 높다는 것을 의미하는 것으로, 충성고객들은 백화점의 기획상품 및 이월상품의 단일가 행사, 한정판매 등의 판촉활동에 의해 다양한 제품을 구매를 함으로써 제품들 간의 밀도가 높게 나올 뿐만 아니라 특정 제품에 대한 선호도가 높아 그 제품을 집중적으로 구매하기 때문에 연결정도 집중도와 연결정도 중심성이 높다고 해석할 수 있다. 반면에 잠재적 이탈고객은 백화점의 판촉활동에 민감하지 않아 구매제품의 종류가 다양하지 않을 뿐 아니라 특정 제품에 대한 선호가 상대적으로 적어 밀도 및 연결정도 집중도가 낮다고 해석할 수 있다.

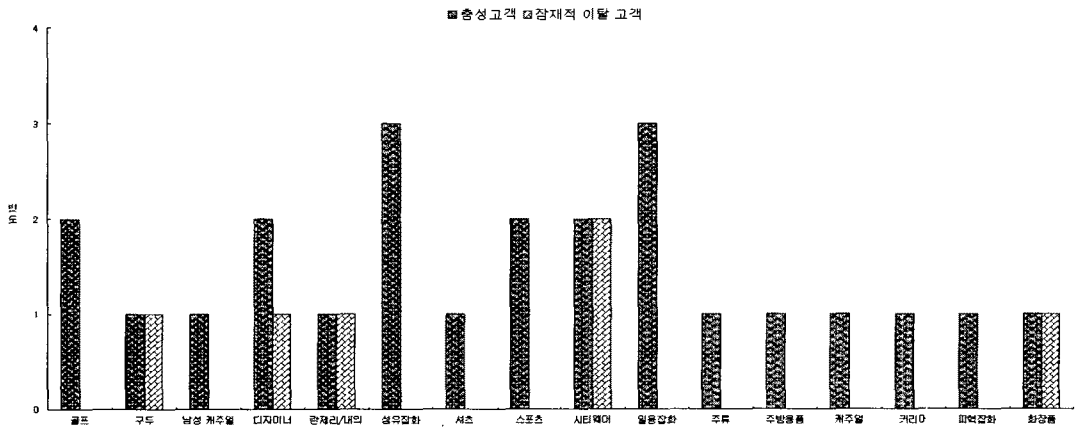
충성고객 그룹과 잠재적 이탈고객 그룹이 선호하는 제품의 차이를 분석하기 위해 연결정도 중심성이 15이상인 제품군을 살펴보면, [그림 7]과 같으며 충성고객이 구매한 제품군은 16개이며, 잠재적 이탈고객이 구매한 제품군은 5개이다. 구두, 디자이너, 란제리/내의, 시티웨어, 화장품은 충성고객과 잠재적 이탈고객이 구매한 제품간에 연결정도 중심성이 높았지만 골프, 남성 캐주얼, 섬유잡화, 셔츠, 스포츠, 일용잡화, 주류, 주방용품, 캐주얼, 커리어, 피혁잡화는 충성고객이 구매한 제품간에 연결정도 중심성이 높았다. 특히 충성고객에 있어 연결정도 중심성이 높은 제품 중 섬유잡화(양말 특정, 스카프 특정, 스타킹 특정)와 일용잡화(위생세제, 뷰티상품, 일

용잡화) 제품군의 빈도가 다른 제품보다 빈도보다 더 높다. 이는 섬유잡화와 일용잡화가 고객을 모으는 집객효과가 있는 미끼상품이라고 해석할 수 있다. 이와 같은 제품의 특성을 이용하면 잠재적 이탈 고객을 충성고객으로 전환시키는 고객 관리 전략을 구축할 수 있다. 예를 들어 생활잡화 등의 기획전 등을 통해 고객 방문을 유도하여 매출을 늘리는 전략을 사용할 수 있다.

4. 결 론

치열한 경쟁환경에서 신규 고객을 획득하는 것보다는 기존 고객을 유지하는 것이 비용, 브랜드 충성도 측면에서 기업에게 유리하다. 많은 연구들이 이탈고객의 특성 분석 및 이탈방지를 위해 인구통계학적 데이터 및 거래데이터 등을 이용하여 통계분석과 데이터마이닝 기법으로 개별 고객의 이탈에 영향을 주는 요인을 분석하였다.

본 연구에서는 고객들이 구매한 제품들 사이의 관계를 분석하기 위하여 실제 L백화점 고객의 구매 데이터를 기반으로 사회네트워크 분석 방법을 이용하여 충성고객의 제품 네트워크와 잠재적 이탈고객의 제품 네트워크를 구성하였다. 또한 각 제품 네트워크의 밀도, 연결정도 집중도 및 연결정도 중심성 분석을 통해 충성고객과 잠재적 이탈고객의 제품



[그림 7] 상대적 연결정도 중심성이 15이상인 제품군

네트워크에 대한 구조적 특징을 도출하였다. 본 연구결과를 간략히 정리하면 다음과 같다.

충성고객의 제품 네트워크와 잠재적 이탈고객의 제품 네트워크에서 아웃디그리 연결정도 중심성이 높은 제품 모두 인디그리 연결정도 중심성이 높았다. 그리고 충성고객의 구매제품 네트워크에서 상대적 연결정도 중심성이 15 이상인 제품의 비중은 잠재적 이탈고객의 구매제품 네트워크보다 약 4배가 높으며, 충성고객의 구매제품 네트워크 밀도 및 연결정도 집중도가 높았다. 이는 충성고객은 백화점에서 다양한 제품을 구매하지만 특정 제품에 대한 선호가 명확하다는 것을 나타낸다. 특히 백화점 상품 중 백화점의 판촉제품 예를 들어 이월상품의 균일가 행사, 한정판매 제품등이 잠재적 이탈고객보다 충성고객에게 집객효과가 크다고 해석할 수 있다.

이 연구는 장바구니 분석을 대체하는 새로운 분석방법을 시도한 연구는 아니며, 오히려 기존의 장바구니 분석을 더 보완하는 분석방법을 제시한 것이 연구의 의의라고 생각한다. 즉, 기존의 장바구니 분석이 의미있는 규칙을 찾는 것이 목적이었다면, 이 연구는 장바구니 분석 결과를 활용하여 제품 네트워크를 구축하면 비교 대상 집단간의 차이를 네트워크의 특성을 이용하여 파악 가능하다는 것이며, 구체적으로 어떻게 차이가 나는지를 실제 데이터를 사용하여 분석하였다는 것이다. 이 연구에서 제시하고 있는 방법론은 새로운 방법은 아니지만, 아직 많이 연구되지 않은 제품 네트워크를 활용하였다는 측면과, 기존의 데이터 마이닝에서 많이 활용되는 장바구니 분석을 활용하여 제품 네트워크를 구축하였으므로, 추후 장바구니 분석을 네트워크 관점에서 확대 적용 가능하다는 측면에서 이 연구의 의의를 찾을 수가 있다.

그러나 본 연구는 고객의 구매 데이터만을 이용하여 제품 네트워크를 구성하였기 때문에 제품 구매에 영향을 줄 수 있는 다른 요인, 예를 들어 상황 데이터를 고려하지 못했을 뿐 아니라 충성고객 250명과 잠재이탈고객 250명을 대상으로 분석하였기

때문에 다양한 고객들의 특성을 반영하지 못했다는 한계점이 있다. 또한 특정기간 동안 충성고객과 잠재적 이탈고객이 구매한 제품에 대하여 네트워크 분석함으로써 시간의 흐름에 따른 제품 네트워크의 변화를 반영하지 못한 한계점이 있다. 따라서 이러한 요인을 고려한 후속 연구가 이루어 진다면 의미 있는 연구가 될 것이다.

참 고 문 헌

- [1] 김용학, 「사회연결망 분석」, 박영사, 2003.
- [2] 김용학, 「사회연결망 이론」, 박영사, 2003.
- [3] 손동원, 「사회 네트워크 분석」, 경문사, 2002.
- [4] 김상용, 송지연, 이기순, "CRM 고객데이터 분석을 통한 이탈고객 연구", 「한국마케팅저널」, 제7권, 제1호(2005), pp.21-42.
- [5] 김재경, 최익영, 김혜경, 김남희, "사회 네트워크 분석을 이용한 충성고객과 이탈고객의 구매 특성 비교 연구", 「경영과학」, 제26권, 제1호(2009), pp.183-196.
- [6] 이희연, 김홍주, "네트워크 분석을 통한 수도권 의 공간구조 변화 1980~2000년", 「국토계획」, 제41권, 제1호(2006), pp.133-150.
- [7] 최창현, "조직의 비공식 연결망에 관한 연구 : 사회연결망 분석의 적용", 「한국사회와 행정연구」, 제17권, 제1호(2006), pp.1-23.
- [8] Agrawal, R. and R. Srikant, Fast algorithms for mining association rules, In Proc. Int. Conf. Very Large Data Bases (VLDB '94), Santiago, Chile, (1994), pp.487-499.
- [9] Barnes, J., "Class and committees in a Norwegian island parish," *Human Relations*, Vol. 7(1954), pp.39-58.
- [10] Berson, A., S. Smith, and K. Thearling, *Building data mining applications for CRM*, McGraw-Hill, 2000.
- [11] Bonacich, P., "Power and Centrality : A Family of Measures," *American Journal of Soci-*

- ology, Vol.92(1987), pp.1170-1182.
- [12] Kauffman, S., *The Origins of Order*, Oxford University Press, Oxford, 1993.
- [13] Lawrence, R.D., G.S. Almasi, V. Kotlyar, M. S. Viveros, and S.S. Duri, "Personalization of Supermarket Product Recommendation," *Data Mining and Knowledge Discovery*, Vol.5, No. 1-2(2001), pp.11-32.
- [14] Ng, K. and H. Liu, "Customer retention via data mining," *Artificial Intelligence Review*, Vol.14, No.6(2000), pp.569-590.
- [15] Raghavan, N., R.M. Bell, and M. Schonlau, "Defection Detection : Using Online Activity Profiles to Predict ISP Customer Vulnerability," *In Proceedings of the sixth International Conference of Knowledge Discovery and Data Mining(2000)*, pp.506-515.
- [16] Sarwar, B., G. Karypis, J.A. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithm," *In Proceedings of The Tenth International World Wide Web Conference*, (2001), pp.285-295.
- [17] Sarwar, B., G. Karypis, J.A. Konstan, and J. Riedl, "Analysis of recommendation algorithms for e-commerce," *In Proceedings of ACM E-commerce 2000 conference*, (2000), pp.158-167.
- [18] Sarwar, B., G. Karypis, J.A. Konstan, and J. Riedl, "Application of Dimensionality Reduction in Recommender System-A case study," *In Proceedings of the ACM web KDD-2000 Workshop*, 2000.
- [19] Song, H., J. Kim, Y. Cho, and S. Kim, "A personalized defection detection and prevention procedure based on the self-organizing map and association rule mining : Applied to online game sit," *Artificial Intelligence Review*, Vol.21, No.2(2004), pp.161-184.
- [20] Trubik, E. and M. Smith, "Developing a model of customer defection in the Australian banking industry," *Managerial Auditing Journal*, Vol.15, No.5(2000), pp.199-208.