

## 사회 네트워크 분석을 이용한 충성고객과 이탈고객의 구매 특성 비교 연구\*

김재경\*\* · †최일영\*\* · 김혜경\*\* · 김남희\*\*

Social Network Analysis to Analyze the Purchase Behavior Of  
Churning Customers and Loyal Customers

Jae Kyeong Kim\*\* · †Il Young Choi\*\* · Hyea Kyeong Kim\*\* · Nam Hee Kim\*\*

### ■ Abstract ■

Customer retention has been a pressing issue for companies to get and maintain the loyal customers in the competing environment. Lots of researchers make effort to seek the characteristics of the churning customers and the loyal customers using the data mining techniques such as decision tree. However, such existing researches don't consider relationships among customers. Social network analysis has been used to search relationships among social entities such as genetics network, traffic network, organization network and so on. In this study, a customer network is proposed to investigate the differences of network characteristics of churning customers and loyal customers. The customer networks are constructed by analyzing the real purchase data collected from a Korean cosmetic provider. We investigated whether the churning customers and the loyal customers have different degree centralities and densities of the customer networks. In addition, we compared products purchased by the churning customers and those by the loyal customers. Our data analysis results indicate that degree centrality and density of the churning customer network are higher than those of the loyal customer network, and the various products are purchased by churning customers rather than by the loyal customers. We expect that the suggested social network analysis is used to as a complementary analysis methodology with existing statistical analysis and data mining analysis.

Keywords : Social Network Analysis, Churning Analysis, Customer Network, Degree Centrality,  
Density

논문접수일 : 2008년 12월 17일      논문수정일 : 2009년 02월 22일      논문제재확정일 : 2009년 03월 02일

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

\*\* 경희대학교 경영대학 경영연구원

† 교신저자

## 1. 서 론

인터넷의 성장으로 인하여 정보의 유통이 많아졌으며, 빠르게 전파되고 있다. 이로 인해 고객들은 그들의 선호에 맞는 제품을 제공하는 기업을 탐색하는데 더 많은 시간을 투자하게 되었으며, 기업들은 고객들을 획득하기 위해 치열한 경쟁을 벌이고 있다. 이러한 경영환경에서 기업은 신규 고객을 획득하는데 많은 비용을 소비하기보다는 구매경험이 있는 기존 고객이 이탈하지 못하도록 노력을 한다면, 더 적은 비용으로 기업 및 제품에 대한 고객의 충성도를 향상시킬 수 있으며 기업의 이익을 높일 수 있다[25, 31].

고객이탈이란 일반적으로 자발적 또는 비자발적으로 고객이 현재의 서비스 이용을 중단하는 것으로 정의되며, 고객의 구매패턴을 바탕으로 이탈고객(Churning customer)을 정의한다. 김상용, 송지연, 이기순[5]의 연구에서는 이탈고객을 분석시점 이후 1년간 구매를 하지 않은 고객으로 정의하였으며, 윤성준[10]의 연구에서는 분석시점 이후 6개월간 구매하지 않은 고객으로 정의하였다. 정리하면, 이탈고객이란 제품도메인에 따라 다르지만 일정기간 동안 구매기록이 없는 고객으로 정의할 수 있다.

고객의 이탈방지 및 특성 분석을 위한 기존의 연구는 성별, 연령, 직업, 거주지역 등과 같은 인구통계학적 데이터와 구매의 최근성, 구매횟수, 구매금액, 구매 제품군, 반품 여부 등과 같은 거래데이터 분석에 기반하고 있다. Berson et al.[15]은 의사결정나무(Decision tree) 기법을 인구통계학적 데이터에 적용하여 고객의 이탈 확률을 분석하였으며, Rehgavan et al.[26] 연구에서는 인터넷 사용기록과 인구통계학적 데이터를 이용한 로지스틱 회귀분석(Logistic regression)을 통해 고객의 이탈 가능성을 예측하였다. Ng and Liu[25]는 데이터베이스와 로그파일을 다양한 데이터마이닝 기법을 통해 분석하여 잠재적으로 이탈 가능성이 있는 고

객을 예측할 수 있는 방법을 제시하였다. 김상용, 송지연, 이기순[5]은 고객이탈여부를 독립변수로 하고 성별, 연령대, 직업, 거주지역, 구매횟수, 구매금액, 구매의 최근성, 주이용채널, 이용채널수, 최초구매제품군, 주이용제품군, 이용제품군 종류 수를 종속변수로 설정한 로지스틱 회귀분석을 통해 유지고객과 이탈고객의 변수별 특성차이를 분석하였으며, 윤성준[10]은 구매금액에 따라 고객 등급을 분류하고, 의사결정나무를 이용하여 고객 등급별 특징을 분석하였다. 그러나 기존 연구는 고객들 간의 관계를 고려하지 않고 개별고객의 이탈에 영향을 주는 변수 분석을 통해 이탈고객의 특징을 파악하였으며 또한 이탈방지 모형을 구축하였다.

고객들 사이의 관계를 분석할 수 있는 방법으로 최근에 사회 네트워크 분석(Social network analysis)에 대한 연구가 증가되고 있다. 사회 네트워크 분석은 의사소통 집단내 개체의 상호작용에 관심을 두고, 개체간 연결 상태 및 연결 구조의 특성을 계량적으로 파악하여 시각적으로 표현하는 분석기법[1, 2, 3]으로 유전 네트워크[16], 교통 네트워크[10], 조직 네트워크[14]등의 구조 분석을 위해 이용되고 있다.

따라서 사회 네트워크 분석을 이용하면 충성고객과 이탈고객의 개별적 속성이 아닌, 충성고객과 이탈고객의 관계적 속성 즉, 연결상태 및 연결 구조의 특성을 계량적으로 분석할 수 있다. 본 연구에서는 고객의 구매 데이터를 이용하여 충성고객 네트워크와 이탈고객 네트워크를 구축하고 사회 네트워크 분석을 통해 충성고객 그룹과 이탈고객 그룹의 연결관계 및 연결구조의 차이를 파악하고자 한다. 또한 각 그룹내의 다른고객과 많은 관계를 유지하고 있는 고객들의 구매 제품 비교를 통해 충성고객 그룹과 이탈고객 그룹의 차이를 파악하고자 한다. 즉 사회 네트워크 분석을 통해 충성고객 그룹과 이탈고객 그룹의 네트워크 구조적 차이 및 구매한 제품 차이 파악이 가능하며 기존의 이탈방지 모형과의 병행을 통해 충성고객과 이탈고객을 더 잘 관리할 것으로 기대된다. 따라서 본 연구의

목적은 고객의 구매 데이터를 활용하여 고객들간의 관계 즉, 충성고객 네트워크와 이탈고객 네트워크를 분석하여 충성고객과 이탈고객의 차이를 파악하는 것으로 국내 F화장품 고객을 대상으로 적용해 보았다.

본 논문의 구성은 다음과 같다. 제 2장에서는 사회네트워크 분석을 살펴보았으며, 제 3장에서는 실제 구매데이터를 이용하여 충성고객 네트워크와 이탈고객 네트워크를 실증적으로 분석하였다. 그리고 제 4장에서는 본 연구의 결론 및 향후 연구방향을 기술하였다.

## 2. 사회 네트워크 분석

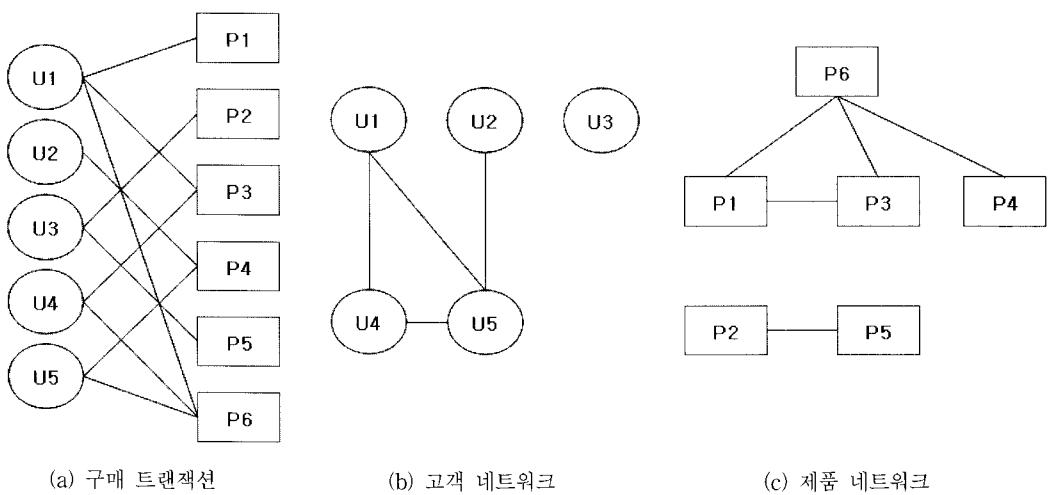
### 2.1 사회 네트워크 분석

사회 네트워크는 Barnes[19]에 의해 처음 사용된 용어로, 초기에 개인적인 인간관계를 기반으로 사람들 사이에 연결한 네트워크이다[3]. 이러한 사회 네트워크는 사회과학뿐만 아니라 점차 경영학, 응용과학 등 다양한 분야에서 응용되고 있으며, 분석 방법으로 다양한 사회 네트워크 분석 기법이 이용되고 있다.

고객의 구매 데이터를 이용하여 사회 네트워크

분석을 하려면, 우선 데이터를 매트릭스 형태로 표현한다. 매트릭스는 행과 열로 구성되며, 행과 열이 만나는 셀에 특정 값을 입력하여 행과 열사이의 관계를 표시한다. 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-제품 매트릭스에서 고객  $i$ 가 제품  $j$ 를 구매하면  $P_{ij} = 1$ , 구매하지 않으면  $P_{ij} = 0$ 으로 표시한다.

고객-제품 매트릭스에서 제품을 구매한 사람들 사이에 직접적인 상호작용의 관계가 없더라도, 관계를 인위적으로 설정하여 고객과 고객 또는 제품과 제품 사이의 관계를 나타낸 네트워크를 준 연결망(Quasi network)이라 한다[1]. <그림 1>은 준 연결망의 예로, 고객의 제품 구매 트랜잭션 (a)를 이용하여 (b), (c)와 같이 고객간의 관계 및 제품간의 관계를 표현하였다. 여기서 (b)를 고객 네트워크라 하며, 고객들이 서로 동일 제품을 1개 이상 구매하였다면 그 고객들은 직접적인 상호작용이 있다고 표현하였다. 일반적으로 고객의 상호 관계를 나타내는 방법은 동일 제품 구매 빈도, 코사인 및 상관계수 계산 등이 있으며 본 연구에서는 상관



<그림 1> 준 연결망 예

계수를 이용하여 고객 네트워크를 구축하였다. (c)는 제품 네트워크로써, 고객이 제품을 동시에 구매하면 그 제품은 서로 상호 관계가 있다고 표현하였다.

본 연구에서는 충성고객과 이탈고객의 네트워크 구조를 파악하기 위해 고객들의 구매데이터를 토대로 노드(Node)와 노드를 연결하는 링크(Link)로 표현하는 사회 네트워크 분석 방법을 사용하였다. 충성고객과 이탈고객은 노드로 표현하고, 고객들의 제품에 대한 선호도가 유사하면 고객들간에 관계가 있다고 판단하여 링크로 연결하였다. 고객들의 제품에 대한 선호도가 유사하다는 것은 고객들의 과거 구매한 제품들 중에서 동일한 제품이 많다는 것을 의미한다.

사회 네트워크 분석방법에서 네트워크 구조를 파악하기 위한 개념으로 중심성(Centrality), 포괄성(Inclusiveness) 및 밀도(Density) 등이 있다[20]. 중심성은 한 행위자가 전체 네트워크에서 중심에 위치하는 정도를 표현하는 지표로 연결중심성(Degree centrality), 근접중심성(Closeness centrality), 매개 중심성(Betweenness centrality)이 있다[1, 2, 3]. 연결중심성은 한 노드에 직접적으로 연결되어 있는 노드의 개수로, 연결된 노드의 수가 많을수록 연결중심성이 높아진다. 노드  $i$ 의 연결중심성( $C_D(i)$ )은 식 (1)과 같이 계산된다.

$$C_D(i) = d(m_i) \quad (1)$$

여기서  $d(m_i)$ 는 노드  $i$ 와 연결된 노드들의 총수이다.

그러나 네트워크간 비교를 위해 절대적 연결중심성을 사용할 경우 네트워크의 노드 개수가 다르면 비교가 불가능하다. 이런 문제점을 해결하기 위해 상대적 연결중심성을 사용하며, 노드  $i$ 의 상대적 연결중심성( $C'_D(i)$ )은 식 (2)와 같이 계산된다.

$$C'_D(i) = C_D(i)/(g-1) \quad (2)$$

여기서  $g$ 는 네트워크에 존재하는 노드들의 총수이다.

근접중심성은 각 노드간의 거리를 근거로 하여 중심성을 측정하는 지표로 한 노드로부터 다른 노드에 도달하기 위해 필요한 최소 단계의 합의 역수로 정의한다. 근접중심성이 높을수록 네트워크에서 중앙에 위치하게 되며, 노드  $i$ 의 근접중심성( $C_C(i)$ )은 식 (3)과 같이 계산된다.

$$C_C(i) = \left[ \sum_{j=1}^n d_{ij} \right]^{-1} \quad (3)$$

여기서  $d_{ij}$ 는 두 노드  $i$ 와  $j$ 를 잇는 가장 짧은 경로거리를 의미한다.

근접중심성은 전체 네트워크 규모에 의해서 각 점 사이에 거리 합의 차이가 발생하기 때문에 네트워크를 비교할 수 없다. 따라서 네트워크간 비교를 위해 상대적 근접중심성이 사용된다. 노드  $i$ 의 상대적 근접중심성( $C'_C(i)$ )은 식 (4)와 같이 계산된다.

$$C'_C(i) = (g-1) C_C(i) \quad (4)$$

여기서  $g$ 는 네트워크에 존재하는 노드들의 총수이다.

매개중심성은 네트워크 내에서 한 노드와 다른 노드들 사이에서 중재자 역할의 정도를 측정하는 방법으로 한 노드가 연결망 내의 다른 노드들 사이의 최다 경로 위에 위치할수록 그 노드의 매개중심성이 높아진다. 노드  $i$ 의 매개중심성은 식 (5)와 같다.

$$C_B(i) = \sum_{j < k} g_{jk}(i)/g_{jk} \quad (5)$$

여기서  $g_{jk}$ 는 네트워크 내 특정 두 노드( $j$ 와  $k$ ) 사이에 존재하는 최단거리경로들의 경우의 숫자이고,

$g_{jk}(i)$ 는 두 노드  $j$ 와  $k$  ( $j \neq k$ ) 사이에 존재하는 노드  $i$ 를 경유하는 횟수를 말한다.

매개중심성 또한 네트워크간 비교를 위해 상대적 매개중심성이 사용된다. 노드  $i$ 의 상대적 매개중심성 ( $C'_B(i)$ )은 식 (6)과 같이 계산된다.

$$C'_B(i) = C_B(i) / [(g-1)(g-2)/2] \quad (6)$$

여기서  $g$ 는 네트워크에 존재하는 노드들의 총수이다.

포괄성은 한 네트워크 내 서로 연결된 행위자들의 수로 정의되며, 포괄성이 높을수록 행위자들 간 관계가 많다는 것을 의미한다. 그러나 포괄성 ( $I_R$ )은 네트워크내에 있는 노드의 수에 영향을 받기 때문에 상대적 비율로서 계산하며, 식 (7)과 같이 계산된다.

$$I_R = (g - i_g) / g \quad (7)$$

여기서  $i_g$ 는 네트워크 내에 연결되어 있지 않고 격리된 노드들의 총수이며,  $g$ 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 (8)과 같다.

$$G = k / [g(g-1)] \quad (8)$$

여기서,  $k$ 는 네트워크 내에 존재하는 링크의 총수이고,  $g$ 는 네트워크에 존재하는 노드들의 총수이다.

본 연구의 목적은 충성고객 네트워크와 이탈고객 네트워크에서 충성고객 그룹과 이탈고객 그룹의 제품 구매 선호도 차이 및 구매 제품의 차이를 분석하는 것이므로, 위에서 제시한 여러 분석 방법

중 연결중심성과 밀도를 사용하기로 한다. 연결중심성은 특정 노드가 다른 노드들과 얼마만큼 관계를 맺고 있는가를 나타내는 노드에 대한 지표이므로 고객들간의 제품 구매 선호도가 얼마나 유사한가를 분석하는데 적합하며 또한 연결중심성이 높은 고객일수록 다른 고객들이 많이 구매하는 제품을 구매하는 고객임을 나타내므로 충성고객 그룹과 이탈고객 그룹의 구매 제품의 차이를 분석하는데 적합하다고 판단된다. 밀도는 네트워크에 있는 노드들간의 응집정도를 나타내는 네트워크에 대한 지표로써, 네트워크의 밀도가 높을수록 동일 제품을 구매한 고객이 많다는 것을 의미하므로 충성고객 네트워크와 이탈고객 네트워크의 선호도 차이를 분석하는데 적합하다고 판단된다. 다른 분석 방법은 실제 데이터 분석에서 충성고객과 이탈고객 그룹간의 뚜렷한 차이가 나타나지 않았으며, 의미 있는 해석을 하는 것이 힘들다고 판단하여, 본 연구에서는 사용하지 않기로 한다.

## 2.2 고객 네트워크 분석

### 2.2.1 고객 네트워크 분석 프로세스

본 연구에서는 제품 구매 데이터를 이용하여 이탈고객 네트워크와 충성고객 네트워크를 구축하고, 네트워크 분석을 통해 이탈고객 네트워크와 충성고객 네트워크 간의 구조적 차이를 분석한다. 즉 이탈고객 네트워크와 충성고객 네트워크의 연결상태, 연결구조 및 응집정도를 분석함으로써, 이탈고객 네트워크와 충성고객 네트워크 간의 차별적 특성을 구하고자 한다.

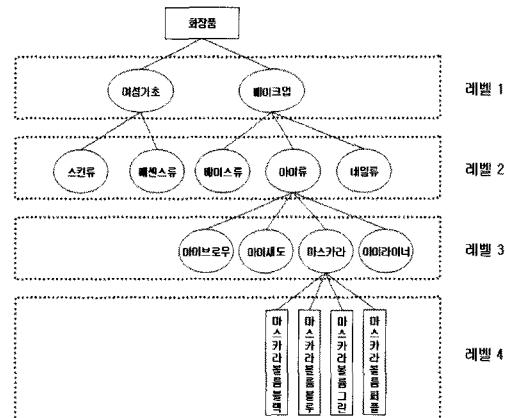
본 연구가 제안하는 고객 네트워크는 유사한 구매 패턴을 가진 고객들 간의 연결망으로, 고객 네트워크 분석 프로세스는 고객 선호 프로파일 생성, 고객 네트워크 구축, 그리고 네트워크 구조 분석의 3단계로 구분된다. 고객 선호 프로파일 생성 단계에서는 이탈고객과 충성고객의 구매 데이터를 이용하여 고객-제품 매트릭스 형태로 이탈고객과 충성고객의 선호를 표현한다. 네트워크 구축 단계에서는

협업 필터링(Collaborative filtering) 기법의 고객들 사이의 유사도를 계산하는 방법을 이용하여 고객 선호 프로파일로부터 특정 고객과 유사한 제품 구매 성향을 가진 이웃고객을 형성한 후 특정 임계치 (Threshold :  $\rho$ ) 이상인 이웃 고객을 링크로 연결 한다. 마지막 단계에서는 고객 네트워크 중심성 분석 및 밀도 분석을 통해 충성 고객과 이탈 고객의 구조적 특징을 파악한다.

### 2.2.2 고객 선호 프로파일 생성

고객 선호 프로파일(Customer preference profile)은 인구통계학적 데이터, 메뉴·컨텐츠·제품 이용, 구매 정보 및 마케팅 피드백 정보 등 다양한 고객 정보를 바탕으로 한 고객 선호를 구조화한 표현이다. 본 연구에서는 Sarwar et al.[27-29]의 연구처럼 고객의 구매 데이터를 이용하여  $i$ 번째 고객이 특정기간에  $j$ 번째 제품을 구매 하였으면 1, 그렇지 않으면 0의 값을 입력한 고객-제품 매트릭스 형태로 고객 선호 프로파일을 생성하였다.

그러나 고객들이 구매하는 제품의 개수는 전체 제품 중 소수이기 때문에 유사한 선호를 가진 고객



〈그림 2〉 제품 계층도 예

을 탐색하는 과정에서 아주 적은 수의 선호 정보가 사용되므로 유사한 선호를 가진 이웃 고객들에 대한 신뢰도가 낮아지는 데이터 희박성(Sparsity)의 문제가 발생한다[21, 28, 29]. 이러한 문제를 해결하기 위해 제시된 방법 중 하나가 제품 계층도(Product hierarchy)이다[21, 22, 24]. 〈그림 2〉는 화장품의 제품 계층도 예를 보여준다.

본 연구에서는 데이터 희박성 문제를 해결하기

〈표 1〉 고객-제품 매트릭스

(a) 충성고객

	여성기초 스킨	여성기초 로션	파우더 팩	클렌징 로션	바디 로션	핸드 케어	파운 데이션	아이 섀도
L <sub>01</sub>	0	1	1	0	0	1	1	0
L <sub>02</sub>	1	0	0	0	0	1	1	0
L <sub>03</sub>	0	0	1	1	0	1	0	1
L <sub>04</sub>	0	1	0	1	0	0	1	0
L <sub>05</sub>	1	1	1	0	0	0	0	1

(b) 이탈고객

	여성기초 스킨	여성기초 로션	파우더 팩	클렌징 로션	바디 로션	핸드 케어	파운 데이션	아이 섀도
D <sub>01</sub>	1	1	1	0	0	1	0	1
D <sub>02</sub>	1	1	0	1	1	1	1	1
D <sub>03</sub>	1	1	0	0	1	1	0	1
D <sub>04</sub>	1	1	0	0	0	0	0	1
D <sub>05</sub>	1	1	1	1	0	1	0	1

위하여 제품계층도의 레벨 3을 이용하여 고객-제품 매트릭스를 생성하였다. <표 1>은 충성고객과 이탈고객의 고객-제품 매트릭스 예이다.

### 2.2.3 고객 네트워크 구축

고객 네트워크는 고객 사이의 관계가 있고 없음을 나타낸 네트워크로, 제품을 구매한 고객들 사이에 직접적인 상호작용이 없어도 동일한 제품을 구매한 고객들은 제품에 대한 구매 성향이 유사하기 때문에 이들 사이를 연결하여 네트워크를 구축할 수 있다. 그러나 동일한 제품을 구매한 고객들 사이에 관계가 있다고 가정하면 모든 고객이 링크로 연결되는 문제점이 있기 때문에, 본 연구에서는 고객 선호 프로파일로부터 제품에 대한 선호가 유사한 고객을 이웃 고객으로 형성한 후 이웃고객 중 유사도가 특정 임계치(Threshold :  $\rho$ ) 이상인 고객들만 관계가 있다고 가정하여 고객들을 링크로 연결하였다.

고객들 사이의 유사도(Similarity)를 측정하는 방법으로는 협업 필터링의 이웃 고객을 형성하는 방법과 마찬가지로 코사인(Cosine)과 피어슨 상관계수(Pearson correlation coefficient)가 있다. Sarwar et al.[28]의 연구에 의하면 피어슨 상관계수와 코사인을 사용하여 고객들간의 유사도를 계산할 때 차이가 없기 때문에 본 연구에서는 식 (9)와 같이 피어슨 상관계수를 이용하여 고객들 사이의 유사도를 계산하였다.

$$\text{sim}(a, b) = \text{corr}_{ab} = \frac{\sum_{j=1}^n (p_{a,j} - \bar{p}_a)(p_{b,j} - \bar{p}_b)}{\left[ \sum_{j=1}^n (p_{a,j} - \bar{p}_a)^2 \sum_{j=1}^n (p_{b,j} - \bar{p}_b)^2 \right]^{1/2}} \quad (9)$$

여기서  $p_{a,j}$  와  $p_{b,j}$ 는 고객  $a$ 와  $b$ 의  $j$  ( $1 \leq j \leq n$ ) 번째 제품에 대한 선호도를 나타내며,  $\bar{p}_a$  와  $\bar{p}_b$ 는 고객  $a$ 와  $b$ 의 평균 선호도를 의미한다.

피어슨 상관계수를 이용하여 계산한 고객들 사

이의 유사도는 -1과 1 사이의 값을 가진다. 유사도 값이 '1'일 경우 두 고객의 구매 제품이 동일하다는 것을 의미하며, '-1'이면 두 고객은 서로 다른 제품을 구매하였다는 것을 의미한다.

<표 2>는 피어슨 상관계수를 이용하여 <표 1>에 대해 고객들 사이의 유사도를 계산한 것이다. <표 1>의 이탈고객을 살펴보면 고객  $D_{01}$ 는 {여성 기초스킨, 여성기초로션, 파우더팩, 핸드케어, 아이섀도}를 구매하였고 {클렌징로션, 바디로션, 파운데이션}은 구매하지 않았다. 또한 고객  $D_{05}$ 는 {여성기초 스킨, 여성기초 로션, 파우더 팩, 핸드케어, 아이섀도}를 구매하였으며 {바디로션, 파운데이션}은 구매하지 않았다. 따라서 고객  $D_{01}$ 과 고객  $D_{05}$ 는 다른 고객들보다 제품 구매 패턴이 유사하기 때문에 상관계수가 0.745로 가장 높게 나타났다.

<표 2> 고객간 유사도

(a) 충성고객

	$L_{01}$	$L_{02}$	$L_{03}$	$L_{04}$	$L_{05}$
$L_{01}$	1				
$L_{02}$	0.258	1			
$L_{03}$	0.000	-0.258	1		
$L_{04}$	0.258	-0.067	-0.258	1	
$L_{05}$	0.000	-0.258	0.000	-0.258	1

(b) 이탈고객

	$D_{01}$	$D_{02}$	$D_{03}$	$D_{04}$	$D_{05}$
$D_{01}$	1				
$D_{02}$	-0.293	1			
$D_{03}$	0.467	0.488	1		
$D_{04}$	0.600	0.293	0.600	1	
$D_{05}$	0.745	-0.218	0.149	0.447	1

고객간 유사도를 계산한 후 제품 구매 패턴이 유사한 고객들 사이의 네트워크를 구축한다. 식 (10)과 같이 고객간 유사도가 특정 임계치,  $\rho$ 이 상인 값을 1로 정의하여 고객들을 링크로 연결하였다.

〈표 3〉 고객 인접 매트릭스

(a) 충성고객

	L <sub>01</sub>	L <sub>02</sub>	L <sub>03</sub>	L <sub>04</sub>	L <sub>05</sub>
L <sub>01</sub>	1				
L <sub>02</sub>	1	1			
L <sub>03</sub>	0	0	1		
L <sub>04</sub>	1	0	0	1	
L <sub>05</sub>	0	0	0	0	1

(b) 이탈고객

	D <sub>01</sub>	D <sub>02</sub>	D <sub>03</sub>	D <sub>04</sub>	D <sub>05</sub>
D <sub>01</sub>	1				
D <sub>02</sub>	0	1			
D <sub>03</sub>	1	1	1		
D <sub>04</sub>	1	1	1	1	
D <sub>05</sub>	1	0	0	1	1

$$R_{a,b} = \begin{cases} 1, & \text{if } sim(a, b) \geq \rho \\ 0, & \text{그렇지 않으면} \end{cases} \quad (10)$$

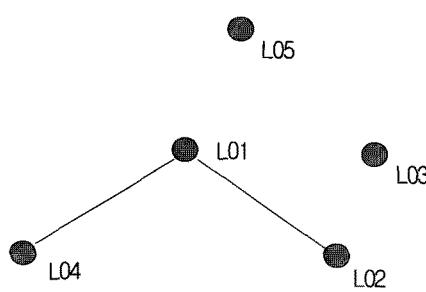
여기서  $R_{a,b}$ 는 고객  $a$ 와 고객  $b$ 의 관계를 나타내고,  $sim(a, b)$ 는 고객  $a$ 와 고객  $b$ 의 고객 유사도를 의미한다.

〈표 3〉은 〈표 2〉의 고객간 유사도 임계치를 0.2로 설정하여 그 이상의 값을 가진 고객들의 인접 매트릭스를 표현하였다. 예를 들어 이탈고객중 고객 D<sub>01</sub>과 다른 고객과의 유사도를 살펴보면  $sim(D_{01}, D_{02}) = -0.293$ ,  $sim(D_{01}, D_{03}) = 0.467$ ,  $sim(D_{01}, D_{04}) = 0.6$ ,  $sim(D_{01}, D_{05}) = 0.745$ 이다. 따라서  $sim(D_{01}, D_{03}) = 0.467$ ,  $sim(D_{01}, D_{04}) = 0.6$ ,  $sim(D_{01}, D_{05}) = 0.745$ 은 0.2보다 크기 때문에 고객 D<sub>03</sub>, D<sub>04</sub>, D<sub>05</sub>는 고객 D<sub>01</sub>에 인접하였다고 할 수 있으므로 고객 인접 매트릭스에 1로 입력하였다.

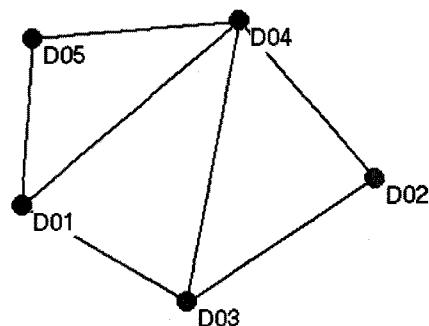
〈그림 3〉은 〈표 3〉을 이용하여 고객 네트워크를 그래프로 표현한 예로 이탈고객과 충성고객은 노드로 표현하였으며, 제품 구매 패턴이 유사한 충성고객 또는 이탈고객은 링크로 연결하였다. 이탈고객 D<sub>03</sub>, D<sub>04</sub>, D<sub>05</sub>와 고객 D<sub>01</sub>의 유사도가 0.2이상이므로 링크로 연결되었으나 이탈고객 D<sub>02</sub>, D<sub>05</sub>는 이탈고객 D<sub>02</sub>과 유사도가 0.2미만이기 때문에 링크로 연결되지 않았다. 마찬가지로 충성고객 중 고객 L<sub>03</sub>와 L<sub>05</sub>은 다른 충성고객과의 유사도가 0.2미만이기 때문에 연결되지 않았다.

#### 2.2.4 네트워크 구조 분석

네트워크 구조 분석은 구축된 충성고객 네트워크와 이탈고객 네트워크의 평균 연결중심성 및 밀도 분석을 통해 각 네트워크의 구조적 특징을 발견하는 단계이다. 여기서 연결중심성은 노드에 대한



(a) 충성고객



(b) 이탈고객

〈그림 3〉 고객 네트워크

지표로서 특정 고객의 연결중심성이 높다는 것은 그 고객과 유사한 제품 구매 패턴을 가진 고객이 많다는 것을 나타낸다. 또한 밀도는 네트워크에 대한 지표로서 네트워크의 밀도가 높다는 것은 그 네트워크에 있는 다수의 고객들의 제품 구매 패턴이 유사하다는 것을 의미한다. 따라서 충성고객 네트워크와 이탈고객 네트워크의 구조 분석을 통해 차이가 발생하면 충성고객과 이탈고객이 선호하는 제품 분석을 통해 충성고객 및 이탈고객에 맞는 마케팅 전략이 가능하게 된다.

<표 4>는 <그림 3>을 이용하여 충성고객 네트워크와 이탈고객 네트워크의 평균 연결중심성 및 밀도를 나타낸 예로 이탈고객 네트워크가 충성고객 네트워크보다 평균 연결중심성과 밀도가 높다. 이는 다수의 이탈고객들이 유사한 제품을 구매하였다는 것을 의미하는 것으로, 앞의 예의 경우 이탈고객들은 충성고객들에 비해 여성기초 화장품을 더 구매했다는 것을 알 수 있다.

<표 4> 고객 네트워크 평균 연결중심성 및 밀도

구 분	충성고객 네트워크	이탈고객 네트워크
노드	5	5
링크	4	16
평균 연결중심성	0.8	3.2
밀도	20%	80%

### 3. 충성고객/이탈고객 네트워크 실증 분석

#### 3.1 실험방법

본 연구에서 제시한 모형에 대한 실증분석을 위하여 사용된 데이터는 국내 화장품 회사 F사의 고객 구매 데이터이다. 2006년 1월부터 12월까지 1년간 F사의 화장품 구매 경험이 있는 사용자를 대상으로 충성고객과 이탈고객으로 구분하였다. 충성고

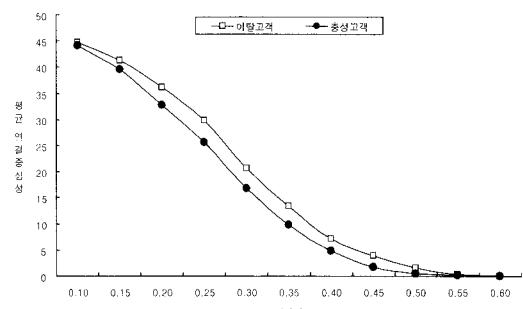
객은 2006년 1월부터 12월까지 매월 구매한 고객으로 정의하였으며, 이탈고객은 1월에서 6월까지 매월 구매하고 7월 이후부터는 구매하지 않은 고객으로 정의하였다. 충성고객과 이탈고객의 제품 구매 선호도를 효과적으로 파악하기 위하여 제품 카테고리 100개 중 최소 8개 이상을 구매한 고객 중 1월부터 6월까지 총 구매금액이 178,000원 이상인 고객을 대상으로 충성고객 50명( $L_1 \sim L_{50}$ ), 이탈고객 50명( $D_1 \sim D_{50}$ )을 무작위 선정하였다. 충성고객 네트워크와 이탈고객 네트워크 규모를 50으로 정한 이유는 시각적으로 네트워크의 구조 차이를 보기 위한 것으로, 실제 분석에서는 네트워크 규모는 다양하게 정할 수 있다.

고객 사이의 피어슨 상관계수를 계산하기 위하여 SPSS 12.0을 사용하였으며, 사회 네트워크 분석을 위해 사용한 소프트웨어는 Ucinet 6.0이다

#### 3.2 충성고객/이탈고객 네트워크 중심성 분석

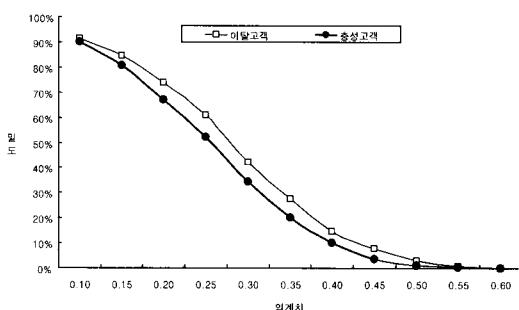
고객의 구매 데이터를 이용하여 고객 네트워크를 구축할 때 고객의 관계 형성을 위한 최적의 임계치는 데이터 특성에 따라 달라진다. 본 연구에서는 이탈고객 네트워크와 충성고객 네트워크의 평균 연결중심성과 밀도의 차이가 가장 크게 나타나는 임계치를 기준으로 분석하였다.

<그림 4>와 <그림 5>는 2006년 1월부터 6월까지 6개월 동안 고객의 구매 데이터를 이용하여 구



<그림 4> 충성고객 네트워크와 이탈고객 네트워크의 임계치별 평균 연결중심성

축한 고객 네트워크에서 임계치별로 충성고객 네트워크와 이탈고객 네트워크의 평균 연결중심성과 밀도 차이를 나타낸 것으로 모든 임계치에서 이탈고객 네트워크의 평균 연결중심성과 밀도가 충성고객 네트워크의 평균 연결중심성 및 밀도보다 더 크게 났으며, 특히 임계치 0.25에서 평균 연결중심성과 밀도의 차이가 가장 크게 나타났다.



〈그림 5〉 충성고객 네트워크와 이탈고객 네트워크의 임계치별 밀도

〈표 5〉는 임계치 0.25에서 충성고객 네트워크와

이탈고객 네트워크의 평균 연결중심성의 차이가 있는지 독립표본 t-검정 결과로 유의확률은 0.02으로 나타났다. 따라서 유의 수준이 0.05보다 작으므로 충성고객 네트워크와 이탈고객 네트워크의 평균 연결중심성의 차이가 통계학적으로 유의한 것으로 나타났다.

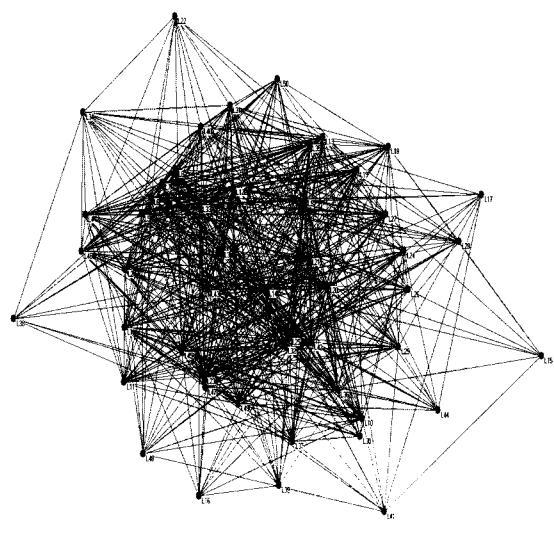
〈표 5〉 임계치 0.25에서 충성고객 네트워크와 이탈고객 네트워크의 독립표본 t-검정 결과

측정치	그룹	평균	t값	유의확률
평균 연결 중심성	충성고객 네트워크	25.6	2.369*	0.02*
	이탈고객 네트워크	29.9		

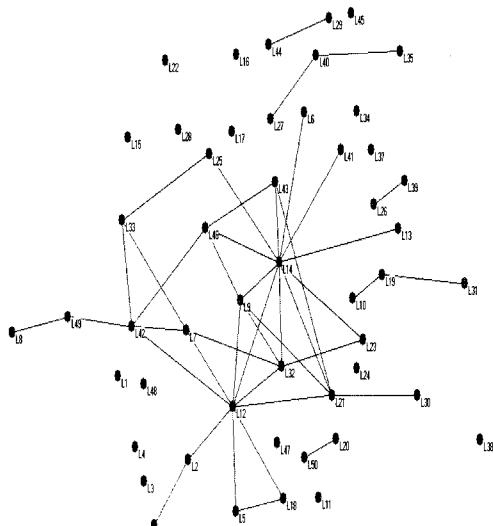
주) \*  $p < 0.05$ .

따라서 본 연구에서는 충성고객과 이탈고객의 차이가 가장 크게 나타나는 임계치 0.25에서 이탈고객 네트워크와 충성고객 네트워크를 분석하였다.

〈그림 6〉과 〈그림 7〉은 임계치 0.25와 임계치

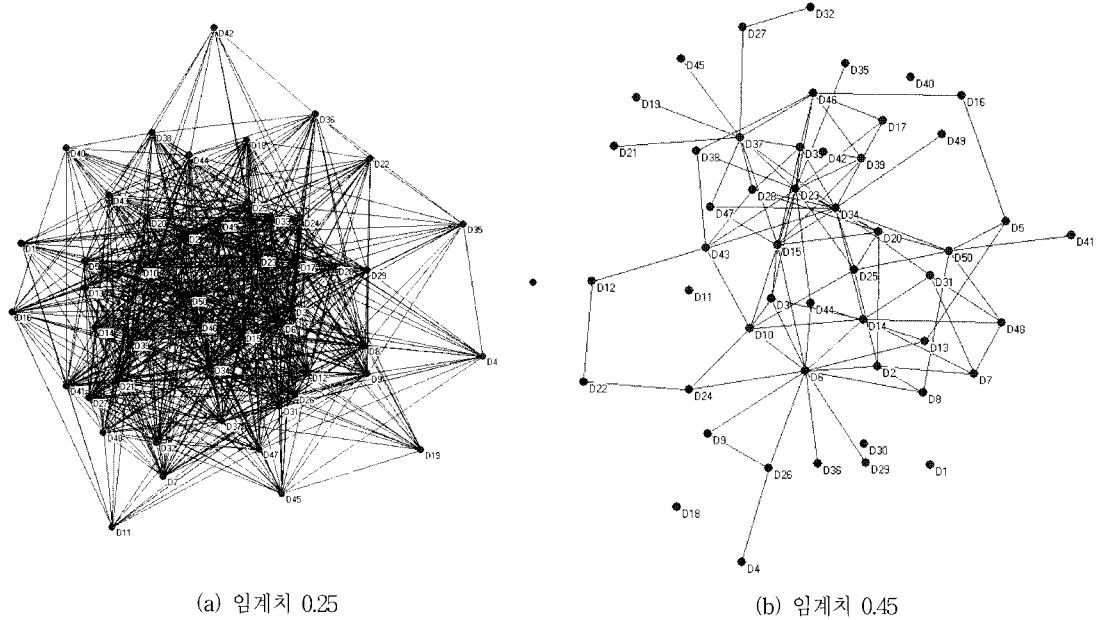


(a) 임계치 0.25



(b) 임계치 0.45

〈그림 6〉 충성고객 네트워크

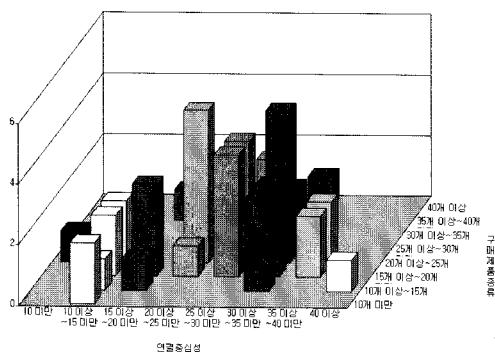


〈그림 7〉 이탈고객 네트워크

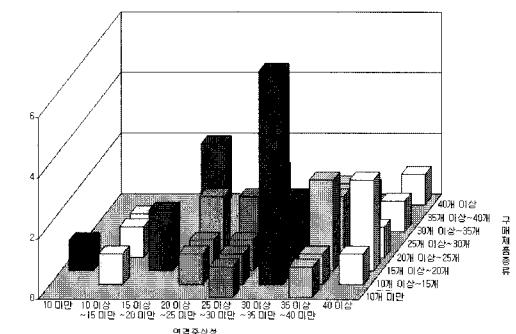
0.45에서 구축한 충성고객 네트워크와 이탈고객 네트워크로, 이탈고객 네트워크가 충성고객 네트워크보다 더 복잡한 네트워크 형태를 나타낸다.

임계치 0.25일 때 충성고객 네트워크와 이탈고객 네트워크의 구조적 특징을 살펴보면 충성고객 네트워크는 구매 제품의 종류가 10개 이상~15개 미만일 때 높은 연결중심성을 보였다. 반면에 이탈고

객 네트워크는 구매 제품 종류의 수에 관계없이 연결 정도가 높았다(〈그림 8〉, 〈그림 9〉 참조). 이러한 구조적 차이를 가져오는 요인은 이탈고객들은 광고, 쿠폰 등의 마케팅 전략으로 인해 다양한 제품을 구매하기 때문에 이탈고객들 간의 연결중심성이 높지만, 충성고객들은 자신이 필요한 품목만을 구매하기 때문에 충성고객들 간에는 연결중심성이 낮다고 해석할 수 있다.

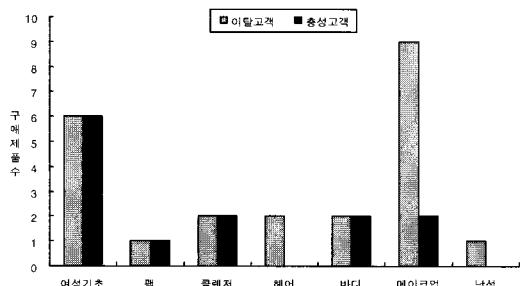


〈그림 8〉 연결중심성과 구매제품종류에 대한 충성고객 분포도



〈그림 9〉 연결중심성과 구매제품종류에 대한 이탈고객 분포도

연결중심성이 높은 충성고객과 이탈고객이 구매한 제품을 파악하기 위해 본 연구에서는 연결중심성이 40이상인 고객들(전체 고객중 80% 이상의 고객과 제품 구매 패턴이 유사한 고객들)을 선정하였다. 연결중심성이 높은 고객들이 구매한 제품은 <그림 10>과 같으며, 충성고객이 구매한 제품은 평균적으로 13개이고 이탈고객이 구매한 제품은 평균적으로 23개이다. 충성고객과 이탈고객 모두 여성기초 화장품, 바디, 팩 제품을 동일하게 구매하였지만, 메이크업, 헤어 및 남성 화장품은 이탈고객이 더 많이 구매하였다. 특히 이탈고객들은 충성고객보다 더 많은 메이크업 제품을 구매함을 알 수 있다. 메이크업 제품은 유행에 민감하여 제품의 라이프 사이클이 짧은 특징을 지니고 있는 화장품으로 이탈고객들은 유행하는 메이크업 제품을 일시적으로 구매하지만 지속적인 제품 구매로 연결되지는 않는다고 해석할 수 있다.



<그림 10> 연결중심성이 40이상인 충성고객과 이탈고객의 구매 제품

충성고객들은 여성기초 화장품 위주로 구매하지만, 이탈고객들은 여성기초 화장품뿐만 아니라 메이크업 제품, 헤어제품, 남성화장품을 구매한다는 것을 알 수 있다. 이와 같은 제품의 판매특성을 이용하면 일시적으로 방문하는 고객 즉, 이탈가능고객을 판별하여 이탈하지 않고 지속적인 구매를 하는 충성고객으로 전환시키는 고객 관리 전략을 구축할 수 있다. 예를 들어 광고에 나오는 메이크업 제품을 매장 전면에 진열하여 신규고객을 유도하는 전략을 사용할 수 있으며, 메이크업, 헤어 또는

남성화장품을 구매하는 고객은 일시적인 구매 가능 고객일 확률이 높으므로 다른 제품의 활인 쿠폰 발송을 통해 고객의 재방문을 유도하여 고객의 이탈 방지 및 충성고객으로 유도하는 전략을 사용할 수 있다.

#### 4. 결 론

지금까지 이탈고객의 특성 분석 및 이탈방지를 위한 연구는 인구통계학적 데이터, 거래데이터 등을 이용하여 통계분석을 통해 개별 고객의 이탈에 영향을 주는 요인을 분석하였다. 또한 데이터마ining 기법을 이용하여 고객의 구매 패턴을 분석하고, 분석된 결과를 바탕으로 잠재적 이탈 고객을 예측하였다.

본 연구에서는 고객들 사이의 관계를 분석하기 위하여 실제 F사의 화장품 구매 데이터를 기반으로 사회 네트워크 분석 방법을 이용하여 충성고객 네트워크와 이탈고객 네트워크를 구축하는 방법을 제시하였다. 또한 각 고객 네트워크의 중심성 분석 및 밀도 분석을 통해 이탈고객과 충성고객의 구조적 특징을 도출하였다. 본 연구결과를 간략히 정리하면 다음과 같다.

충성고객 네트워크는 구매 제품의 종류가 10개 이상~15개 미만일 때 높은 연결중심성을 보였으나 이탈고객 네트워크는 구매 제품 종류의 수에 관계없이 연결중심성이 높았다. 또한 이탈고객 네트워크가 충성고객 네트워크보다 밀도가 높게 나타났다. 이는 이탈 고객들은 광고, 쿠폰 등의 마케팅 활동과 관계된 특정 제품을 구매함으로써 고객간 유사성이 높은 반면에, 충성 고객들은 자신에게 필요한 품목 중심으로 구매함으로써 고객간 유사성이 낮다고 해석할 수 있다. 또한 충성고객과 이탈고객의 구매 제품을 분석한 결과 이탈고객은 충성고객보다 메이크업 제품을 더 많이 구매함을 알 수 있었다.

그러나 본 연구는 고객의 구매 데이터만을 이용하여 고객 네트워크를 구성하였기 때문에 제품 구

때에 영향을 줄 수 있는 다른 요인, 예를 들어 고객의 인구통계학적 특성 데이터를 고려하지 못했을 뿐 아니라 충성고객 50명과 이탈고객 50명을 대상으로 분석하였기 때문에 다양한 고객들의 특성을 반영하지 못했다는 한계점이 있다. 또한 고객을 충성고객과 이탈고객으로 이분하게 구분하여 분석 함으로써 시간의 흐름에 따른 고객네트워크의 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] 윤영수, 채승병, 「복잡계 개론」, 삼성경제연구소, 2007.
- [5] 김상용, 송지연, 이기순, "CRM 고객 데이터 분석을 통한 이탈고객 연구", 「한국마케팅저널」, 제7권, 제1호(2005), pp.21-42.
- [6] 김재경, 채경희, 송희석, "SOM을 이용한 온라인 게임 제공업체의 고객이탈방지 방법론", 「경영과학」, 제21권, 제3호(2004), pp.85-99.
- [7] 박수진, 박길순, 김서연, "여대생의 베이크업에 대한 행동 및 의식 조사연구-대전, 충남지역을 중심으로", 「한국패션뷰티학회지」, 제4권, 제4호(2006), pp.87-99.
- [8] 변현수, 김진화, "인터넷 쇼핑몰 사용자가 구축하는 연결망 구조분석을 통한 구매행동 모델링", 「대한경영학회지」 제20권, 제5호(2007), pp.2069-2091.
- [9] 송기룡, 노성호, 이재광, 최일영, 김재경, "고객의 동적 선호 탐색을 위한 순차패턴 분석 : 주더페이스샵 사례", 「Information Systems Review」, 제10권, 제2호(2008), pp.195-209.
- [10] 윤성준, "데이터마이닝 기법을 통한 백화점의 고객이탈예측 모형 연구", 「한국마케팅저널」, 제6권, 제4호(2005), pp.45-72.
- [11] 이희연, 김홍주, "네트워크 분석을 통한 수도권의 공간구조 변화 1980~2000년", 「국토계획」, 제41권, 제1호(2006), pp.133-150.
- [12] 장덕진, "연결망과 사회이론", 「사회와 이론」 제4집(2003), pp.11-61.
- [13] 조영빈, 조윤호, "구매순서를 고려한 개선된 협업필터링 방법론", 「한국지능정보시스템학회 논문지」, 제13권, 제2호(2007), pp.69-80.
- [14] 최창현, "조직의 비공식 연결망에 관한 연구 : 사회연결망 분석의 적용", 「한국사회와 행정 연구」, 제17권, 제1호(2006), pp.1-23.
- [15] A Berson, S Smith, K Thearling, *Building data mining applications for CRM*. McGraw-Hill, 2000.
- [16] Kauffman, S., *The Origins of Order*, Oxford University Press, Oxford, 1993.
- [17] Breiman, L., J.H. Friedman, R.A. Olshen, and C.J. Stone, *Classification and Regression Trees*, Wadsworth and Brooks/Cole Advanced Books and Software, Pacific Grove, California 1984.
- [18] Anderson, C., A. Hellervik, and K. Lindgren, "A Spatial Network Explanation for a Hierarchy of Urban Power Laws", *Physica A : Statistical Mechanics and its Applications*, Vol.345(2004), pp.227-244.
- [19] Barnes, J. "Class and committees in a Norwegian island parish," *Human Relations*, Vol.7(1954), pp.39-58.
- [20] Bonacich, P. "Power and Centrality : A Family of Measures," *American Journal of Sociology*, Vol.92(1987), pp.1170-1182.
- [21] 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 (2004), pp.233-246.

- [22] Cho, Y.H., J.K. Kim, and S.H. Kim, "A personalized Recommender System based on Web Usage Mining and Decision Tree Induction," *Expert Systems with Applications*, Vol.23(2002), pp.329-342.
- [23] Kim, Y., H. Song, and S. Kim, "A dynamic procedure for defection detection and prevention based on SOM and a Markov chain," *International Conference of Korea Intelligent Information Systems Society*, (2003), pp.141-148.
- [24] 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.
- [25] Ng, K. and H. Liu, "Customer retention via data mining," *Artificial Intelligence Review*, Vol.14, No.6(2000), pp.569-590.
- [26] 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.
- [27] 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.
- [28] 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.
- [29] Sarwar, B., Karypis, G., Konstan, J.A. and Riedl, J., "Application of Dimensionality Reduction in Recommender System A case study," *In Proceedings of the ACM web KDD-2000 Workshop*, 2000.
- [30] 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.
- [31] 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.
- [32] Yeo, A., K. Smith, R. Willis, and M. Brooks, "Modeling the effect of premium changes on motor insurance customer retention rates using neural networks," *Lecture notes in computer science*, (2001), pp.390-399.