

사회연결망 : 신규고객 추천문제의 새로운 접근법

박종학

동양공업전문대학 인터넷비즈니스과
(jhpark@dongyang.ac.kr)

조윤희

국민대학교 경영학부
(www4u@kookmin.ac.kr)

김재경

경희대학교 경영대학 & 경영연구원
(jaek@khu.ac.kr)

.....

협업필터링은 상품을 추천하고자 하는 고객과 유사한 구매 행태를 보이는 고객들의 구매 정보를 반영하여 추천대상 고객이 아직 구매하지 않은 상품에 대한 선호도를 예측한 후 선호도가 높을 것으로 예측되는 상품을 추천해주는 시스템이다. 그러나 신규고객의 경우에는 과거 구매 이력의 부재로 선호도를 예측할 수 없어 추천이 어렵게 되는 신규고객 추천문제가 발생하게 된다. 이러한 신규고객 추천문제를 해결하기 위해 기존에 제시되었던 방법들은 추천의 정확도가 낮거나, 추천에 필요한 정보 획득이 어렵거나, 추천 전에 고객이 능동적으로 질의에 응답해야 하는 부담이 있는 등의 문제로 인하여 그 실효성이 매우 낮다. 따라서 기존의 신규고객 추천 방법의 한계를 극복할 수 있는 새로운 접근방법의 필요성이 대두되고 있다.

본 연구에서는 사회네트워크 분석에서 관계·구조적 특성을 분석하기 위해 널리 활용 되고 있는 중심성 개념을 협업필터링에 적용하여 신규고객의 이웃고객을 찾고 그 이웃고객들의 구매정보를 이용하여 신규고객에게 상품을 추천하는 방법을 제시한다. 추천 프로세스는 구매 유사도 분석, 고객 네트워크 구성, 이웃고객 형성, 신규고객 상품추천 단계로 구성된다. 제시한 추천방법의 성능을 평가하기 위하여 국내 유명 백화점 중의 하나인 H백화점의 고객 구매 데이터를 사용하여 실험하였다. 실험 결과로부터 제시한 추천방법이 기존의 신규고객 추천방법들과 비교하여 추천의 정확도는 높으면서도, 구매정보 외에 인구통계정보 등과 같은 추가 정보가 필요하지 않으며, 추천 전에 고객이 능동적으로 질의에 응답할 필요가 없는 새로운 방법임을 알 수 있었다.

.....

논문접수일 : 2009년 02월 05일 논문수정일 : 2009년 02월 27일 게재확정일 : 2009년 03월 17일 교신저자 : 조윤희

1. 서론

인터넷의 급속한 확장으로 인하여 인터넷 쇼핑 물들이 질적으로나 수적으로 고속 성장함에 따라 이들 간의 경쟁이 치열해지게 되었고 이들 기업들에게는 필연적으로 경쟁력에 있어 상대적 우위 확보 수단에 대한 요구가 증가하게 되었다. 고객의 입장에서 본다면 인터넷 쇼핑물이 증가함에 따라 더 많은 상품들 그리고 더 많은 공급자들이 선택

가능해진 것은 사실이다. 그러나 이와 함께 고객 역시 이들 수많은 상품들 중 어떤 상품이 정말 고객 자신의 요구에 정확히 부합되는 상품인지를 찾고 또한 평가하기 위해 더 많은 노력을 기울여야 하는 부담 역시 발생하게 되었다(Cho and Kim, 2004). 이와 같은 이유로 기업은 고객별로 차별화된 일대일 마케팅(One-to-One Marketing)과 고객의 입장에서 고객을 이해하고 고객과의 관계를 강화시켜 나가는 고객관계관리(Customer Rela-

* 이 연구는 한국학술진흥재단(과제번호 : 20071183)의 ‘복잡계 네트워크 이론에 기반한 추천방법 연구’ 과제로부터 지원을 받아 수행되었음.

tionship Management, CRM) 전략 등을 사용함으로써 기업의 경쟁력 강화에 주력하고 있다. CRM의 여러 분야 중 전자상거래시스템에서 중요한 이슈로 떠오르는 것이 바로 고객이 좋아하는 제품이나 서비스를 추천해주는 추천시스템(Recommender System)이다. 추천시스템은 자동화된 정보 필터링 기술을 이용하여 고객의 취향에 맞는 상품을 추천해주는 시스템으로, Amazon.com을 비롯한 국내외 유수의 인터넷 쇼핑몰에서 널리 활용되고 있다. 추천시스템에서 가장 중요한 것은 고객의 선호도를 정확히 분석하고 예측하여 고객이 원하는 상품을 적시에 추천할 수 있는 능력이다. 이를 위하여 수많은 추천방법들이 개발되어 왔는데, 이 중에서 협업필터링(Collaborative Filtering) (Sarwar, et al., 2000; Adomavicius and Tuzhilin, 2005; 김재경 등, 2005)이 가장 성공적인 상품추천 방법으로 알려져 있으며 많이 이용되고 있다.

CF는 상품을 추천하고자 하는 고객과 유사한 구매 행태를 보이는 고객(이웃고객이라고 함)들의 구매 정보를 반영하여 추천대상 고객이 아직 구매하지 않은 상품에 대한 선호도를 예측한 후 선호도가 높을 것으로 예측되는 상품을 추천해주는 시스템이다. 그러나 신규고객의 경우에는 과거 구매 이력이 없어 선호도를 예측할 수 없으므로 추천이 어렵게 되는 신규고객 추천문제(Huang, et al., 2004; Adomavicius and Tuzhilin, 2005)가 발생하게 된다. 이러한 신규고객 추천문제를 해결하기 위한 기존 연구로는 가장 많이 팔린 상품을 신규고객에게 추천하는 베스트셀러 기반 추천방법(Sarwar, et al., 2000), 고객 속성을 분류 또는 규칙화하여 추천에 활용하는 인구통계정보 기반 추천방법(Krulwich, 1997; Aggarwal, et al., 1998), 최소의 질의로 고객 선호도를 파악하기 위한 최소 질의대상 상품 결정방법(Schein, et al., 2002; Yu, et

al., 2004; Park, et al., 2006) 등이 있다. 하지만 이러한 방법들은 추천의 정확도가 낮거나, 추천에 필요한 정보 획득이 어렵거나, 추천 전에 고객이 능동적으로 질의에 응답해야 하는 부담이 있는 등의 문제로 인하여 그 실효성이 매우 낮다. 따라서 기존의 신규고객 추천 방법의 한계를 극복할 수 있는 새로운 접근방법이 모색되어야 한다.

최근에 사람들 사이의 사회적 관계를 분석할 수 있는 방법으로 사회연결망 또는 사회네트워크 분석(Social Network Analysis)이 많이 연구되고 있다. 사회네트워크 분석은 의사소통 집단 내 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. CF에서는 고객들 간의 상품 선호도 또는 구매 연관성을 분석한 후 유사한 고객들을 묶어 상품을 추천하는 관계를 형성하는데, 이 관계를 그래프로 나타내어 네트워크를 형성하면 이 네트워크는 사회네트워크가 된다(Ryu, et al., 2006). 즉, CF 추천시스템은 고객의 구매데이터를 분석하여 사회네트워크를 인위적으로 생성한 후에 링크로 연결된 고객들(이웃 고객)의 구매정보를 이용하여 특정 고객에게 상품을 추천할 수 있게 만든 일종의 사회네트워크 시스템이라고 할 수 있다.

일반적으로 사회네트워크에 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. 예를 들어, 어느 개인이 특정 사회집단(사회네트워크)에 지금 막 소속되고자

한다고 가정하자. 그 사람은 해당 사회집단으로부터 재화, 용역, 정보 등 필요한 것을 얻고자 할 때 아직은 자신을 도와 줄 인적 관계, 소위 인맥이 제대로 형성되어 있지 않기 때문에 그 사회집단에서 가장 인맥이 넓은 사람들에게 도움을 청하는 것이 일반적인 사회현상이고 또한 가장 합리적인 의사 결정이다. 또한 그 과정에서 그들 간의 인맥(네트워크 연결)도 자연스럽게 형성된다. CF도 하나의 사회네트워크이기 때문에 이웃고객으로 가장 많이 참조되는 고객을 인맥이 가장 넓은 고객으로, 지금 막 CF 기반의 사회네트워크에 진입하려는 고객을 신규고객으로 해석할 수 있다. 따라서 미래에 신규고객은 현재 이웃고객으로 가장 많이 참조되고 있는 고객들에게 연결될 가능성, 다시 말해 그들과의 구매 유사 확률이 높기 때문에, 사회네트워크 분석을 이용하여 그들을 찾고 그들이 구매한 상품들을 신규고객에게 추천한다면 CF에서의 신규고객 추천문제를 추가적인 정보의 필요나 선호도 입력에 대한 부담 없이 해결 할 수 있게 된다.

이와 같은 아이디어를 기반으로 본 연구에서는 사회네트워크 분석에서 관계 및 구조적 특성을 분석하기 위해 널리 활용되고 있는 중심성(Centrality) 분석기법을 CF에 적용하여 신규고객의 잠재 이웃고객을 찾고 그 이웃고객들의 구매정보를 이용하여 신규고객에게 상품을 추천하는 새로운 방법을 제시한다. 또한 제시한 추천방법을 국내 유명 백화점 구매 정보를 이용하여 제시한 방법의 특징을 살펴보고 기존의 추천방법과 비교함으로써 그 성능을 검증하고자 한다.

2. 관련 연구

2.1 신규고객 추천문제

추천시스템은 통계적 기법과 지식탐사기술(Kno-

wledge 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(CF)이 가장 성공적인 추천 기법으로 알려져 있으며 인터넷 쇼핑몰 등에서 널리 활용되고 있다. CF 기반 추천시스템은 상품을 추천하고자 하는 고객과 취향이 유사한 고객들의 의견을 반영하여 추천대상 고객이 아직 구매하지 않은 상품에 대한 선호도를 예측한 후 선호도가 높을 것으로 예측되는 상품을 추천해주는 시스템이다. 일반적으로 CF 기반 추천 프로세스는 크게 입력 데이터 구성, 이웃 집단 탐색, 추천 상품 결정 단계로 구성된다(Sarwar, et al., 2000).

그러나, 인터넷 비즈니스에서 관리하는 상품과 고객의 수가 급속히 증가하면서 협업필터링에 기반한 상품추천시스템은 다음과 같은 세 가지 심각한 문제점을 노출시키고 있다(Sarwar, et al., 2000; Adomavicius and Tuzhilin, 2005; 조운호 등, 2004). 첫째는 희박성(Sparsity)의 문제이다. CF 기반 추천시스템은 고객의 선호도 데이터를 많이 확보할수록 추천의 정확도가 높아진다. 그러나 인터넷 비즈니스의 성장과 함께 인터넷 쇼핑몰서 취급하는 상품이 많아짐에 따라 고객의 직접 평가나 구매정보 분석을 통하여 수집되는 선호도 데이터가 존재하지 않은 상품의 개수가 상대적으로 많아진다. 따라서 고객-상품 행렬은 희박 행렬(Sparse Matrix)일 수밖에 없으며, 유사집단을 탐색하는 과정에서 아주 적은 수의 선호도 데이터를 사용하므로 고객들 간의 유사도 측정 시 신뢰성이 떨어지게 된다. 이러한 현상은 결국 추천결과의 정확도를 떨어뜨리게 하는 주요인으로 작용한다. 둘째는

추천 알고리즘의 확장성(Scalability) 문제이다. CF 기반 상품추천 프로세스에서 유사 집단 탐색 과정은 일종의 lazy learning과 유사하다. 따라서 고객과 상품의 수가 증가함에 따라 고차원화된 고객-상품 행렬로부터 유사 고객을 찾기 위한 연산량은 기하급수적으로 늘어 날 수밖에 없기 때문에 실시간 추천을 목적으로 하는 추천시스템은 심각한 시스템 확장성 문제에 직면하게 된다. 셋째는 신규고객의 상품추천 문제이다. CF 기반 추천시스템에서 고객이 상품을 추천 받으려면 그 동안 구매한 구매데이터나 상품에 대한 선호도 정보가 필요하다. 그러나 인터넷 쇼핑물에 처음 방문하는 신규고객에게는 상품에 대한 선호도 정보나 상품에 대한 구매이력이 전혀 없으므로, CF 방법으로는 신규고객에게 상품을 추천하는 것이 근본적으로 어렵다.

이와 같은 CF의 한계점을 극복하기 위하여 많은 연구자들은 추천 대상 고객이 선호하는 상품과 유사한 특성을 가진 상품을 추천하는 내용기반 필터링(Content Based Filtering)을 CF와 결합하여 사용하였다. 이러한 하이브리드 추천방식은 위에서 논의한 첫 두 가지 문제(희박성 및 확장성 문제)에 대해서는 보다 향상된 결과를 가져왔다(Melville, et al., 2001; Cho and Kim, 2004; Kim and Yum, 2005; 조윤호 등, 2004; 김재경 등, 2005). 그러나 신규고객 추천문제는 하이브리드 추천 방식으로도 해결이 불가능하다. 따라서 CF의 신규고객 추천 문제를 보완하기 위해 베스트셀러 기반 추천방법(Sarwar, et al., 2000), 인구통계정보 기반 추천방법(Krulwich, 1997; Aggarwal, et al., 1998), 최소 질의대상 상품 결정방법(Schein, et al., 2002; Yu, et al., 2004; Park, et al., 2006) 등이 현재까지 주로 활용되어 왔다.

베스트셀러 기반 추천방법은 인터넷 쇼핑물, 전

자상거래 등에서 가장 많이 팔린 순서로 상품을 정렬하여 신규고객에게 상품을 추천하는 방법이다. 이 방법은 간단하다는 점과 함께 신규고객이 쇼핑물에 접속하면 바로 상품을 추천해줄 수 있는 장점이 있어 실무에서는 많이 활용되고 있다. 그러나 모든 신규고객에게 동일한 상품을 추천하게 되어 개인별 성향을 고려하는 개인화된(Personalized) 추천을 할 수 없으며, 많이 팔린 상품들이 특정 상품 군에 속하는 경우가 많아 다양한 종류의 상품을 추천하지 못하게 되어 추천 정확도가 낮은 것으로 알려져 있다.

인구통계정보 기반 추천방법은 신규고객이 인터넷 쇼핑물에 가입할 때 입력한 인구통계적 정보를 상품을 추천하는데 이용하는데, 신규고객과 비슷한 인구통계적 정보를 가진 기존고객이 구매한 상품을 신규고객에게 추천하는 방식(Krulwich, 1997)과 기존고객의 인구통계정보와 구매기록과의 연관성을 규칙으로 이 규칙을 신규고객에게 적용하여 추천하는 방식(Aggarwal, et al., 1998)이 있다. 이 방법은 가입과 동시에 신규고객에게 개인화된 상품을 추천해 줄 수 있다는 점뿐만 아니라 획득한 인구통계적 정보를 향후에 다른 용도로도 사용할 수 있다는 장점이 있다. 하지만 수집한 정보의 양이 부족하거나 정보의 신뢰성이 결여될 수 있는 등 정확한 인구통계정보를 획득하기 어렵고, 정보처리 시간이 많이 소요되며, 인구통계정보와 구매결과 간의 상호 관련성이 높지 않기 때문에 추천 정확도도 높지 않은 것으로 보고되고 있다.

최소 질의대상 상품 결정방법은 특정 상품 목록에 대한 신규고객의 선호도를 추천시스템에 입력하라고 질의하는 것을 가정하고, 선호도 입력을 최소로 하면서도 신규고객의 선호도를 충분히 파악할 수 있게 하는 최소의 질의대상 상품목록을 결정하는 방법이다. 질의대상 상품목록을 결정하기

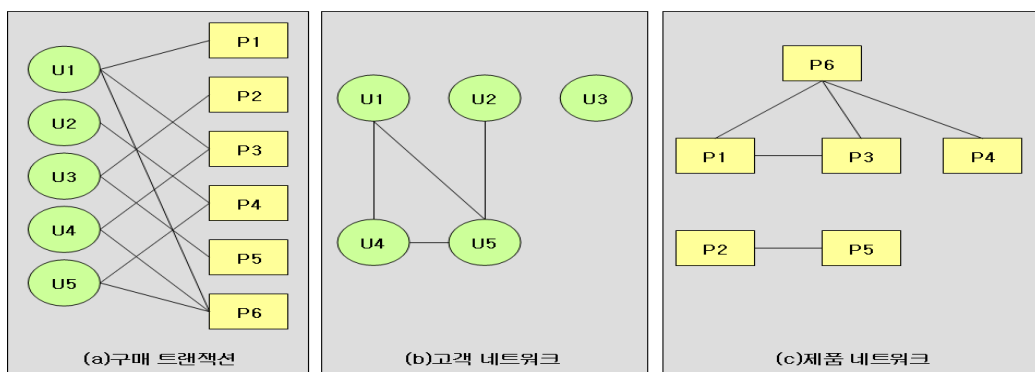
위해 상품의 대중성(Popularity), 상품의 엔트로피(Entropy), 고객 개인화 등에 기반을 둔 알고리즘들이 연구되고 있다. 그러나 이 방법은 추천 전에 고객들로 하여금 추천시스템에서 요구하는 상품에 대한 선호도를 묻는 질의에 능동적으로 응답해야 하는 부담을 주며, 질의에 따라 입력된 선호도 정보가 부정확할 경우 추천 정확도가 떨어지는 단점이 있다.

2.2 사회네트워크 분석

사회네트워크는 Barnes(1954)에 의해 처음 사용된 용어로, 개인적인 인간관계가 확산되어 형성된 사람들 사이의 연결된 네트워크이다(손동원, 2002). 이러한 사회 네트워크는 사회과학뿐만 아니라 경영학, 응용과학 등 다양한 분야에서 응용되고 있으며, 분석 방법으로 사회 네트워크 분석이 이용되고 있다. 일반적으로 사회 네트워크 분석에서는 데이터를 매트릭스 형태로 표현한다. 매트릭스는 행과 열로 구성되며, 행과 열이 만나는 셀에 특정 값을 입력하여 행과 열 사이의 관계를 표시한다. 여기서 행과 열에 같은 개체가 배열되는 것을 1원(1 Mode)자료라 하며, 다른 개체가 배열되는

경우를 2원(2 Mode) 자료라 한다. 개체와 개체 간의 관계가 존재하면 1, 존재하지 않으면 0으로 입력하는 것이 관계를 표현하는 기본적인 방법이다. 예를 들어, 고객-제품 매트릭스(2원 자료)에서는 고객 i 가 제품 j 를 구매하면 셀 p_{ij} 를 1로, 구매하지 않으면 0으로 표시한다.

고객-제품 매트릭스에서처럼 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(김용학, 2003). 아래 <그림 1>은 준 연결망의 예로, 고객의 구매 트랜잭션 (a)를 이용하여 (b), (c)와 같이 고객간의 관계 및 제품간의 관계를 표현한 것이다. 여기서 (b)를 고객 네트워크라 하며, 고객들이 서로 동일 제품을 1개 이상 구매하였다면 그 고객들은 직접적인 상호작용이 있다고 표현한 것이다. 일반적으로 고객의 상호 관계를 나타내는 방법으로는 동일 제품 구매 빈도, 코사인 및 상관계수 계산 등이 있으며 본 연구에서는 상관계수를 이용하여 고객 네트워크를 구축하였다. (c)는 제품 네트워크로서, 고객이 함께 구매한 제품들 간에는 서로 상호 관계가 있다고 표현한 것이다.



<그림 1> 준 연결망의 예

사회네트워크 분석기법은 여러 개의 기법들로 구성되어 있고 지금도 새로운 기법이 만들어지고 있는 일종의 구성체이다. 사회네트워크 분석방법에서 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)이 있다(Bonacich, 1987; 손동원, 2002).

- **연결정도 중심성** : 한 노드에 직접적으로 연결되어 있는 노드들의 합으로 얻어진다. 한 노드에 얼마나 많은 다른 노드들이 관계를 맺고 있는가를 통해서 그 노드가 중심에 위치하는 정도를 계량화한 것이다. 연결정도 중심성은 한 노드의 포인트 중심성을 측정하는 방법이다. 예를 들면 연결정도 중심성은 나와 직접적으로 관계를 맺고 있는 사람 또는 회사가 얼마나 많은가를 가지고 중심성을 표시한다. 연결된 노드의 수가 많을수록 연결정도 중심성이 높아진다.
- **근접 중심성** : 각 노드간의 거리를 근거리 하여 중심성을 측정하는 방법으로 연결정도 중심성과는 달리 직접적으로 연결된 점뿐만 아니라 네트워크 내 간접적으로 연결된 모든 노드들 간의 거리를 합산하여 중심성을 측정한다는 특징을 가진다. 예를 들어 내가 속해있는 어떤 집단에서 내가 집단 내에 있는 모든 사람들과 만난다고 가정하자. 나는 집단 내에 있는 모든 사람들을 알지는 못한다. 하지만 내가 아는 사람들을 거쳐서 가면 모든 사람들

을 알 수 있게 된다. 이때 가장 짧은 경로로 모든 사람을 알게 되는 것이 근접 중심성이다. 근접 중심성은 각 노드간의 거리를 근거리 하여 중심성을 측정하는 지표로 한 노드로부터 다른 노드에 도달하려면 필요한 최소 단계의 합으로 정의한다. 근접 중심성이 높을수록 네트워크에서 중앙에 위치하게 된다.

- **매개 중심성** : 네트워크 내에서 한 노드가 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 매개 중심성이 높다. 한 노드가 연결망 내의 다른 노드들 사이의 최다 경로 위에 위치하면 할수록 그 노드의 매개 중심성이 높아진다.

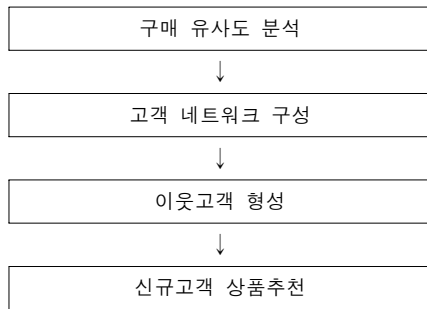
CF에서는 추천 대상 고객과 직접 연결된 이웃 고객과의 관계만 고려하고, 대상 고객의 이웃의 이웃 고객까지의 관계는 고려하지 않는다. 따라서 네트워크 안의 모든 고객과의 관계를 고려하는 근접 중심성과 어떤 그룹과 관계를 맺고 있는지를 더 중요하게 보는 매개 중심성은 CF에 기반을 둔 본 연구에서 활용하기에는 적합하지 않다. 따라서 본 연구에서는 위의 세 가지 중심성 중에 연결정도 중심성만을 적용하여 신규고객의 이웃고객을 찾고자 한다.

3. 사회연결망 분석 기반 신규고객 추천 방법

3.1 신규고객 추천 프로세스

<그림 2>는 사회네트워크 분석을 활용한 신규

고객 추천 방법의 개괄적인 절차를 보여준다. 모두 4단계로 추천 프로세스가 구성되는데, 첫 번째는 ‘구매 유사도 분석’ 단계로 기존 상품을 구매했던 고객들의 구매데이터를 분석하는 단계이다. 기존 고객의 구매데이터를 가지고 각각 고객들의 유사도를 구해 유사한 구매패턴을 보이는 고객들을 찾는다. 두 번째는 ‘고객 네트워크 구성’ 단계로 분석된 구매데이터를 가지고 사회네트워크를 구성하는 단계이다. 유사한 구매패턴을 가진 고객들을 링크(link)로 연결하여 하나의 그래프를 만드는 단계이다. 세 번째는 ‘이웃고객 형성’ 단계로 구성된 사회네트워크에서 연결정도 중심성(이하 중심성)을 구해 중심성이 높은 고객을 찾아 신규고객의 이웃고객이 되게 하는 단계이다. 마지막으로 ‘신규고객 상품 추천’ 단계로 신규고객의 이웃고객이 된 고객들이 구매했던 상품을 신규고객에게 추천하는 단계이다.



<그림 2> 신규고객 추천 프로세스

3.2 단계 1 : 구매 유사도 분석

단계 1에서는 기존 고객들의 구매성향을 분석하여 고객 간 유사도를 구하는데, 구매성향을 분석하는 단계와 유사도를 구하는 단계로 나누어진다. 구매성향을 분석하기 위해, 구매 트랜잭션 DB로부터 기존 M 명의 고객이 L 개의 상품 중 어떤 것을 구매

했는지를 파악하고 그 결과를 식 (1)과 같은 고객-상품 매트릭스 $\mathbf{P} = (p_{ij})$ 로 표현한다. 즉, \mathbf{P} 는 고객의 상품 구매여부에 따라 이진 값을 갖는 매트릭스이다. 만약 한 고객이 같은 상품을 여러 개 구입하더라도 구매는 한 번 일어난 것으로 간주하여 1로 표현한다.

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

고객-상품 매트릭스가 만들어지면, 식 (2)와 같이 피어슨 상관계수를 이용하여 고객들 사이의 유사도를 계산한다. 피어슨 상관계수를 이용하여 계산한 고객들 사이의 유사도는 -1에서 +1 사이의 값을 갖는다. 유사도 값이 '1'일 경우 두 고객의 구매 상품이 동일하다는 것을 의미하며, '1'이면 두 고객은 서로 상이한 상품을 구매하였다는 것을 의미한다. 식 (2)에서 p_{ak} 와 p_{bk} 는 기존고객 a 와 b 의 상품 k 에 대한 구매여부를 나타내며 \bar{p}_a , \bar{p}_b 는 고객 a 와 b 의 평균선택도율을 의미한다.

$$\text{sim}(a, b) = \text{corr}(a, b) = \frac{\sum_{k=1}^L (p_{ak} - \bar{p}_a)(p_{bk} - \bar{p}_b)}{\sqrt{\sum_{k=1}^L (p_{ak} - \bar{p}_a)^2 \sum_{k=1}^L (p_{bk} - \bar{p}_b)^2}} \quad (2)$$

3.3 단계 2 : 고객 네트워크 구성

단계 2에서는 단계 1에서 계산한 유사도로 고객 네트워크를 구성하는 단계이다. 고객간 유사도를 계산한 후 제품 구매 패턴이 유사한 고객들 사이의 네트워크를 구축한다. 네트워크는 동일한 제품을 구매한 고객들 사이에 관계가 있다고 가정하면

거의 모든 고객이 링크로 연결되는 문제점이 있기 때문에 고객간 유사도가 특정 임계치 ρ 이상인 값을 1로 정의하여 고객들을 링크로 연결하였다. 고객 a 와 b 간의 연결정도 $link(a, b)$ 는 식 (3)과 같이 정의된다.

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

이와 같이 형성된 고객네트워크에서는 비슷한 구매패턴을 가진 이웃이 링크로 연결된다. 한 고객에게 연결된 링크가 많다는 것은 이웃이 많다는 것이고, 이웃이 많다는 것은 이웃들과 유사한 상품을 많이 구매했다는 것을 의미한다.

3.4 단계 3 : 이웃고객 형성

단계 3에서는 신규고객의 이웃고객을 정하는 단계이다. 이 연구에서는 단계 2에서 구성된 고객네트워크에서 중심성을 구하여 중심성이 높은 고객을 신규고객의 이웃고객으로 선택한다. 즉, 중심성을 사용하여 중심성 값이 높은 상위 K 명을 뽑아 신규고객의 이웃고객(H)으로 하여 이웃고객이 구매한 상품을 신규고객에게 추천하게 하는 것이다. 신규고객에게 중심성 값이 높은 고객을 이웃으로 선정한 이유로는, 중심성 값이 높은 고객은 다른 고객들보다 상대적으로 많은 사람들과 유사한 제품을 구매하기 때문에 CF에서 좋은 이웃고객이 될 확률이 높기 때문이다. CF에서 좋은 이웃고객이란 이웃고객으로 선택된 고객들이 추천한 제품을 추천 받은 고객이 구매하는 경우가 많은 고객을 말한다. 상품을 많이 구매한 사람이 CF에서 좋은 이웃고객이 될 확률이 높은 것은 아니다. 상품을 많이 구매했다더라도 구매한 상품이 다른 고객들

과 동떨어진 상품을 많이 구매했다면 유사도가 떨어지므로 중심성 값도 떨어지게 된다. 이 연구에서는 이와 같이 네트워크 특성에 기반한 휴리스틱 방법을 제시하고, 제시한 방법의 성능을 추천방법의 정확도를 사용하여 실제 데이터를 이용하여 분석하였다.

고객 c 의 중심성 $cen(c)$ 은 고객 c 에 직접 연결되어 있는 고객 노드의 수를 의미하며, 식 (4)와 같이 표현될 수 있다.

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

3.5 단계 4 : 신규고객 상품추천

단계 4에서는 단계 3에서 선정한 K 명의 중심성이 높은 고객들이 과거에 구매한 상품들을 찾아 각 상품의 구매가능성을 계산하여 구매가능성이 높은 상위 $N(Top-N)$ 개의 상품을 신규고객에게 추천한다. 신규 고객 c 의 j 번째 상품에 대한 구매가능성점수(purchase likelihood score) $pls(c, j)$ 는 식 (5)와 같이 표현된다.

$$pls(c, j) = \frac{\sum_{i \in H} p_{ij} \times cen(i)}{\sum_{i \in H} cen(i)} \quad (5)$$

식 (5)에서 p_{ij} 는 고객 i 의 j 번째 상품에 대한 구매부를 나타내며 $cen(i)$ 는 고객 i 의 중심성값, H 는 단계 3에서 선택된 신규고객의 이웃 집합을 의미한다. pls 값을 계산하여 값이 큰 순으로 정렬하여 상위 N 개의 상품을 신규고객에게 추천한다. CF에서 pls 값을 사용하여 추천 상품을 정하는 이유는 pls 값이 높다는 것은 이웃 고객들의 구매빈도가 많다는 것이고 이것은 신규고객도 구매할 확

률이 높다는 것을 의미한다. 식 (5)에서 기존 CF와
의 차이점은 신규고객은 구매기록이 없으므로, 다
른 고객과의 유사도를 계산할 수 없기 때문에, 유
사도 대신에 중심성 값이 높은 기존 고객들을 신
규고객의 이웃고객으로 사용하는 것이다.

4. 신규고객 추천 예제

이 장에서는 제 3장에서 설명한 1단계에서 4단
계까지의 신규고객 추천방법을 구체적인 예제를
통해 설명하기로 한다. 이 예제에서는 임의의 인터
넷 쇼핑몰에서 점퍼, DMB, 구두, 에어컨, 손목시
계, 전기밥솥, 코트, 헤어드라이어, 반팔티셔츠, 가
방과 같은 10개의 상품을 팔고 있다고 가정한다.
이 인터넷 쇼핑몰에서 상품을 구매하였던 고객은
총 10명이며, 이들은 서로 아무 연관성도 없고 자
신 외 다른 구매자가 누구인지 모르는 상황이라고
가정한다. 또한 본 연구에서 제시하는 방법을 따라
추천시스템을 구축하였고 D일부터 신규고객에 대
한 추천을 시작한다고 하자.

4.1 구매 유사도 분석

인터넷 쇼핑몰에서 판매하는 상품은 점퍼(P₁),
DMB(P₂), 구두(P₃), 에어컨(P₄), 손목시계(P₅), 전
기밥솥(P₆), 코트(P₇), 헤어드라이어(P₈), 반팔티셔
츠(P₉), 가방(P₁₀)이다. 10명의 기존고객 C = {C₁,
C₂, ..., C₁₀}가 10개의 상품 P = {P₁, P₂, ..., P₁₀}를
D-1일까지 구매한 기록을 <표 1>과 같이 정리
하였다. 기존고객들이 가장 많이 구매한 상품은
DMB(P₂), 밥솥(P₆)이고 가장 많은 상품을 구매한
기존고객은 8개의 상품을 구매한 C₃이고 가장 적
은 상품을 구매한 기존고객은 1개의 상품을 구매
한 C₇이다.

<표 1> 기존고객의 구매상품

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

<표 1>에 식 (1)을 대입하여 구매는 '1', 비구매
는 '0'으로 나타내면, <표 2>와 같은 고객-상품 매
트릭스를 만들 수 있다. 또한 <표 2>와 식 (2)를
사용하여 기존고객 10명의 유사도를 구하면 <표
3>과 같은 결과를 얻게 된다.

가장 높은 유사도를 보인 고객은 유사도가
0.857인 C₅와 C₉로 C₅는 점퍼(P₁), DMB(P₂), 에어
컨(P₄), 전기밥솥(P₆), 코트(P₇), 헤어드라이어(P₈),
가방(P₁₀)를 구매하였고 C₉는 점퍼(P₁), DMB(P₂),
전기밥솥(P₆), 코트(P₇), 헤어드라이어(P₈), 가방
(P₁₀)를 구매하였다. C₅와 C₉가 같이 구매한 상품
은 점퍼(P₁), DMB(P₂), 전기밥솥(P₆), 코트(P₇), 헤
어드라이어(P₈), 가방(P₁₀)으로 에어컨(P₄)만 다르
다. 가장 낮은 유사도를 보인 고객은 유사도가
-0.764인 C₅와 C₆, 그리고, C₆와 C₈이다.

4.2 고객 네트워크 구성

이전 단계에서 계산된 유사도 값이 임계치 ρ 보

<표 2> 고객-상품 매트릭스

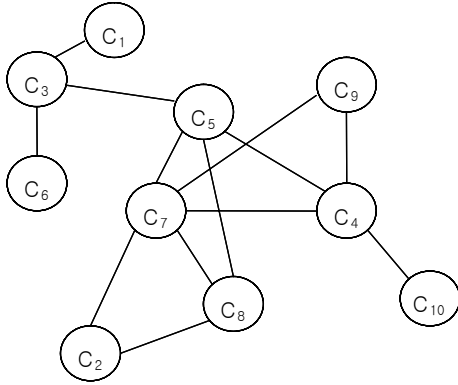
	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇	P ₈	P ₉	P ₁₀
C ₁	0	1	1	1	0	0	0	1	0	0
C ₂	0	0	0	0	1	1	1	0	0	0
C ₃	1	1	1	1	0	1	1	1	1	0
C ₄	1	1	0	0	1	1	0	1	0	1
C ₅	1	1	0	1	0	1	1	1	0	1
C ₆	0	0	1	0	0	0	0	0	1	0
C ₇	0	0	0	0	0	1	0	0	0	0
C ₈	0	1	0	1	0	1	1	0	0	1
C ₉	1	1	0	0	0	1	1	1	0	1
C ₁₀	1	1	0	1	1	0	0	0	0	0

다 크면 두 고객의 관계를 링크로 연결하고 유사도가 ρ 보다 낮으면 두 고객의 관계를 형성하지 않는다. ρ 은 도메인 및 구매데이터의 특성에 따라 달라지는데, 추천시스템 관리자가 실험이나 경험을 통해 적절히 선택해야 한다. 이 예제에서는 ρ 를 0.3으로 설정하였고, 이에 따라 기존 고객들 간의 네트워크를 구성하면 <그림 3>과 같은 결과를 얻는다. 가장 많은 상품을 구매한 사람은 C₈이고 가

장 적은 물건을 구매한 사람은 C₇이다. 그러나 <그림 3>에서 보면 C₈에는 C₂, C₅, C₇ 세 명이고 C₇에는 C₂, C₄, C₅, C₈, C₉ 다섯 명으로 C₈보다 C₇과 연결된 사람이 많다. 이것은 관계가 형성될 때 많은 상품을 구매한 고객이 많은 고객과 이웃하는 것이 아니고, 여러 고객과의 유사한 상품을 많이 구매한 고객이 많은 고객과 이웃이 될 수 있음을 의미한다.

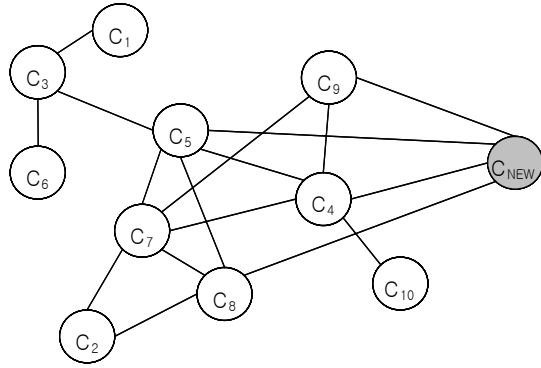
<표 3> 기존고객의 유사도

	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈	C ₉	C ₁₀
C ₁		-0.53452	0.408248	-0.16667	0.089087	0.102062	-0.27217	0.000000	-0.10000	0.166667
C ₂			-0.21822	-0.03563	-0.04762	-0.32733	0.509175	0.218218	0.089087	-0.08909
C ₃				-0.40825	0.218218	0.25000	0.166667	0.000000	0.102062	-0.10206
C ₄					0.356348	-0.61237	0.272166	0.163299	0.583333	0.25000
C ₅						-0.76376	0.218218	0.654654	0.801784	0.008909
C ₆							-0.16667	-0.50000	-0.61237	-0.40825
C ₇								0.333333	0.272166	-0.27217
C ₈									0.408248	0.00000
C ₉										-0.16667
C ₁₀										



<그림 3> 기존고객들로 구성된 고객 네트워크

을 링크로 연결한 고객 네트워크를 보여준다.



<그림 4> 신규고객이 추가된 고객 네트워크

4.3 이웃고객 형성

본 예제에서는 D-1일까지 발생한 구매정보를 가지고 고객 네트워크를 구성하였다. 구성된 고객 네트워크부터 D일에 인터넷 쇼핑몰에서 처음 상품을 구매하려는 신규고객에게 상품을 추천할 중심성 값이 가장 높은 기존 고객을 찾아 신규고객의 이웃고객으로 정하려고 한다. 본 예제에서 이웃의 수 K 는 5명으로 설정하였다. 식 (4)를 이용하여 중심성 값을 계산하면 다음 <표 4>와 같다.

신규고객의 이웃고객으로 추천된 고객은 중심성 값이 높은 상위 5명의 고객 {C₄, C₅, C₇, C₈, C₉}이다. <그림 4>는 신규고객과 5명의 이웃고객들

4.4 신규고객 상품추천

중심성 값이 높은 상위 5명의 고객이 구매한 상품중 pls 값이 높은 N 개의 상품을 신규고객에게 추천한다. 이 예제에서는 추천상품수 N 은 3개로 설정하였으며, pls 값을 식 (5)를 이용하여 구하면 <표 5>와 같다.

pls 값이 가장 높은 상품은 P₆으로 이웃고객 5명 모두 구매하여 1이 나왔고 가장 작은 값 0이 나온 P₃, P₉는 5명 중 아무도 구매하지 않았다는 것을 의미한다. 결론적으로 신규고객에게 추천할 상품은 밥솥(P₆), DMB(P₂), 가방(P₁₀)이다.

<표 4> 각 고객의 중심성 값

	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈	C ₉	C ₁₀
중심성	1	2	3	4	5	1	5	4	4	1

<표 5> 상품별 pls 값

	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇	P ₈	P ₉	P ₁₀
pls	0.591	0.773	0	0.409	0.182	1	0.591	0.591	0	0.773

5. 성능 평가

5.1 실험 데이터

본 연구에서 제시한 신규고객 추천방법의 성능을 평가하기 위하여 국내 유명 백화점 중의 하나인 H백화점의 고객 구매 데이터를 사용하여 실험하였다. 실험에서 사용한 데이터의 상세 내역은 다음과 같다.

- 구매 기간 : 2000년 5월 1일부터 2001년 4월 30일까지 1년간
- 구매 고객 수 : 3,960명
- 구매 상품 수 : 96개
- 구매 건수 : 16,664건

5.2 실험 목적 및 방법

본 연구에서 제시한 추천방법의 특성을 파악하고 기존의 추천방법과의 성능을 비교하기 위하여 다음과 같은 연구질문을 가지고 실험하였다.

첫째, 유사도를 다르게 설정하면 이에 따라 추천 정확도가 달라지는가? 유사도 임계치 ρ 을 지나치게 낮게 설정하면 고객 노드 간의 링크가 너무 많이 생겨 중심 고객을 구분하는 변별력이 떨어지게 되고, 반대로 너무 높게 설정할 경우에는 고객네트워크의 생성 자체가 어려워질 것이다. 유사도에 따라 추천 정확도가 다른지를 보기 위하여 실험에서는 유사도 임계치 ρ 을 0.2, 0.4, 0.6, 0.8로 변화시켜 가면서 추천 정확도를 측정하였다.

둘째, 이웃고객의 수를 다르게 설정하면 이에 따라 추천 정확도가 달라지는가? 일반적으로 CF에서 이웃고객의 수가 너무 적거나 많으면 추천 정확도가 떨어진다고 알려져 있다(Sarwar et al., 2000; Cho and Kim, 2004). 제시한 방법에서도 이

와 같은 현상이 나타나는지를 보기 위하여 이웃고객의 수 K 를 20명에서 200명까지 변화시켜 가면서 추천 정확도를 측정하였다.

셋째, 사전학습기간을 다르게 설정하면 이에 따라 추천 정확도가 달라지는가? 기존 고객의 구매 데이터를 많이 사용하면 할 수록 보다 많은 정보(고객 선호도, 구매 내역 등)를 반영하여 보다 정확한 고객네트워크가 구축될 가능성이 높을 것이다. 이는 CF에서 구매 정보가 많아짐에 따라 희박성이 개선되어 추천 정확도가 향상되는 것과 유사하다. 따라서 신규고객 추천 시점 이전의 구매 데이터 분석 기간(사전학습기간)을 3개월 단위로 증가시켜 가면서 추천 정확도의 변화를 확인하였다.

넷째, 기존의 추천방법보다 추천 정확도가 높은가? 기존 신규고객 추천방법과의 성능을 비교하기 위하여 베스트셀러 기반 추천시스템을 구현하였고, 제안 추천방법의 속성별(유사도, 사전학습기간, 이웃고객의 수) 추천 정확도와 서로 비교하였다.

많은 기존연구에서는 추천 정확도를 평가하기 위한 척도로서 재현율(recall)과 정확율(Precision)을 자주 사용하여 왔다(Sarwar et al., 2000; Ahn, 2008). 그러나 이들 평가척도는 계산이 간단하고 쉽게 이해되는 장점은 있으나, 추천 집합의 크기가 커지면 재현율은 올라가고 정확율을 떨어지는 상반된 결과를 보인다(Sarwar et al., 2000). 그래서 이들 척도를 식 (6)과 같이 동일한 가중치로 결합한 F1 measure(Sarwar et al., 2000; Cho and Kim, 2004; 조운호 등, 2004)가 최근의 추천시스템 관련 연구에서 자주 사용되고 있는데 본 연구에서도 F1을 사용하여 제안하는 추천방법의 추천 정확도를 측정하였다.

$$F1 = \frac{Recall \times Precision}{(Recall + Precision)/2} \quad (6)$$

실험을 위한 유사도, 사전학습기간, 이웃고객의 수가 결정되면, 사전학습기간(제 4절 예제에서의 D-1 일까지) 동안 구매한 고객들을 대상으로 고객네트워크를 형성하고 사전학습기간 이후(예제에서의 D일부터) 매일마다 당일 처음으로 구매한 고객들을 식별하여 그들이 구매한 상품과 고객네트워크로부터 추천되는 10개의 추천상품(즉, $N = 10$)을 비교하여 추천 정확도를 계산하였다. 이 과정에서 사전학습기간 이후 매일 마다 당일의 신규고객을 포함하여 이제까지 구매한 기록이 있는 모든 고객들을 대상으로 고객네트워크를 새로 형성하고, 이렇게 생성된 고객네트워크를 이용하여 익일에 신규고객 추천이 이루어지도록 하였다. 이러한 실험 방식은 실제로 추천시스템이 구현되어 신규고객에게 서비스될 때와 동일한 환경 하에서 실험을 수행하기 위한 것으로서, 실제 서비스 환경과 실험 환경이 상이하지만 연구의 편의상 기준에 많이 이용되어 왔던 학습집단(Training Set)과 검증집단(Test Set)으로 분할하여 실험하는 방식(Sarwar, et al., 2000; Huang, et al., 2004; Ahn, 2008)이 갖는 문제점을 개선한 새로운 접근법이다.

5.3 실험 결과 및 분석

유사도, 사전학습기간, 이웃고객의 수를 변화시켜 가면서 실험한 결과를 도식하면 <그림 5>와 같다. <그림 5>에서 볼 수 있는 것처럼 유사도에 따라 추천 정확도가 달라지는 것을 알 수 있다. 유사도가 낮을 때($\rho = 0.2$, $\rho = 0.4$)는 베스트셀러 기반 추천방법보다 거의 모든 경우에 제안 추천방법의 추천 정확도가 낮은 것으로 나타났고, 유사도가 증가하면서 추천 정확도가 높아지지만 적정 수준($\rho = 0.6$)보다 높아 질 때는 추천 정확도가 다시 낮아지는 것을 알 수 있다. 이는 유사도를 지나치게

낮게 설정하면 고객 노드 간의 링크가 너무 많이 생겨 중심 노드를 구분하는 변별력이 떨어지게 되고, 반대로 너무 높게 설정할 경우에는 고객네트워크의 생성 자체가 어려워져 추천 정확도가 떨어질 것이라는 우리의 예상과 일치한다. ANOVA 검정 결과 통계적으로도 유사도에 따른 추천 정확도의 차이가 유의한 것으로 검증되었다($F = 90.26$, $p < 0.01$).

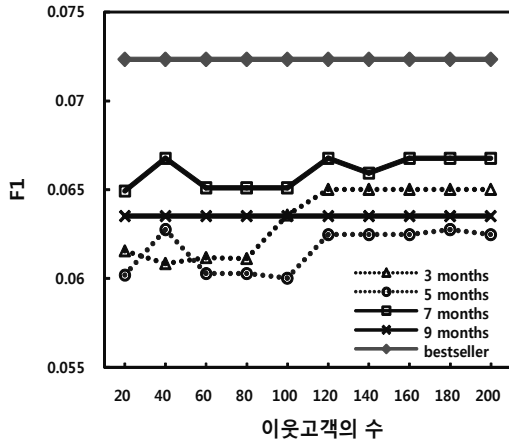
<그림 5>는 사전학습기간에 따라 추천 정확도가 달라지는 것을 보여준다($F = 2.6$, $p < 0.1$). 전체적으로 볼 때 사전학습기간이 길면 길수록 보다 고객 선호도를 반영한 고객네트워크가 형성될 것이라는 우리의 예상과는 달리 기간에 따라 그 차이가 유의할 뿐 사전학습기간이 길수록 추천 정확도가 반드시 향상되지는 않는다는 결과를 얻었다. 그러나, 유사도가 높을 때($\rho = 0.6$, $\rho = 0.8$)는 사전학습기간이 길수록 추천 정확도가 향상되는 것을 볼 수 있는데, 이는 사전학습기간과 유사도가 서로 함께 추천 정확도에 유의한 영향을 준다는 것을 의미한다($F = 6.39$, $p < 0.01$). 이런 결과로부터 유사도가 낮을 때는 짧은 기간 동안에 이미 많은 링크가 형성되어(즉, 고객네트워크가 성숙되어) 사전학습기간이 길어져도 별 영향을 미치지 못하지만, 유사도가 높을 때는 시간이 지남에 따라 천천히 고객네트워크가 성숙되어 추천 정확도가 높아진다고 유추해 볼 수 있다.

실험결과로부터 이웃고객의 수는 추천 정확도에 유의한 영향을 미치지 않는 것으로 판명되었다. 다만, 유의도가 0.6인 경우에는 이웃고객의 수가 많아짐에 따라 추천 정확도가 증가했다가 감소하는 것을 볼 수 있는데, 이는 앞서 설명한 CF의 연구 결과와 일치한다. 유사도가 0.8일 때는 거의 대부분의 차트가 이웃고객의 수가 증가할수록 추천 정확도가 증가하고 있는데, 이런 현상은 유사도가

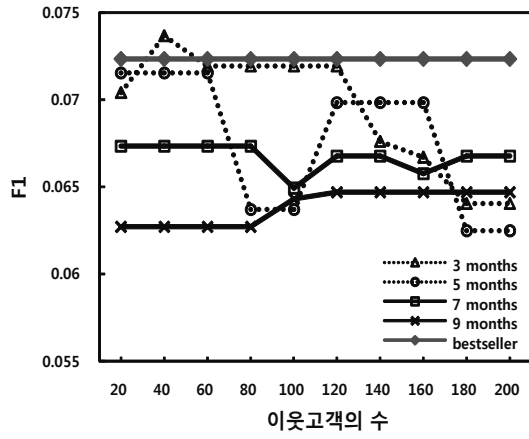
높아 링크가 적게 연결된 고객네트워크에서는 중심성 값이 동일한 고객이 많아 뚜렷한 이웃고객을 식별하기가 어려우나, 이웃고객의 수가 증가하면서 이점이 상쇄되고 있음을 나타낸다.

마지막으로 베스트셀러 기반 추천방법과 추천 정확도를 비교해 보았다. 베스트셀러 기반 추천방법도 사전학습기간의 변화에 따라 추천 정확도가 조금씩 달라지는데, <그림 5>의 베스트셀러 기반

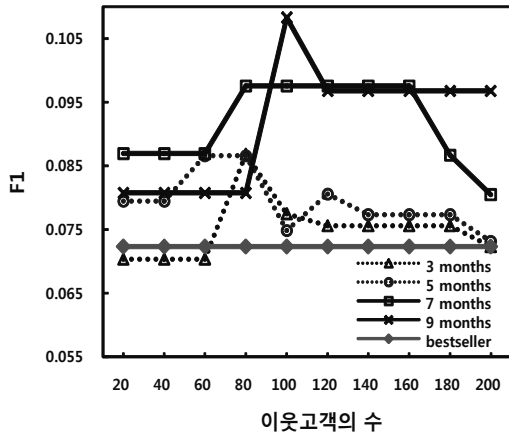
추천방법의 정확도는 이 중 최대의 치를 나타낸 것이다. 적정 수준의 유사도($p = 0.6$) 하에서는 이웃고객의 수나 사전학습기간에 따른 추천 정확도를 평균하여 계산하면 추천 정확도가 28.5% 향상됨을 알 수 있다. 특히, 이웃고객의 수가 100명이 고 사전학습기간이 9개월 일 경우에는 무려 49.7%나 추천 성능을 향상시킨다. 이 결과와 앞에서 논의한 결과로부터, 적절한 유사도와 이웃고객의 수



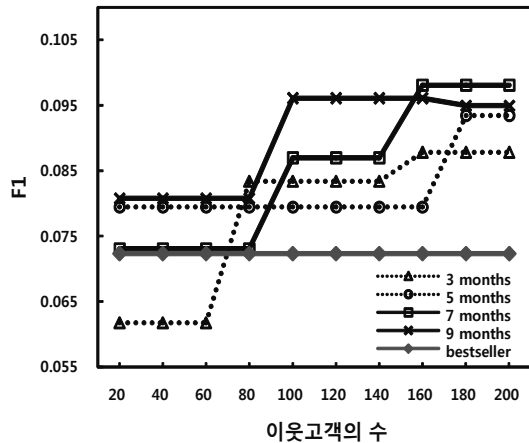
(a) 유사도가 0.2 일 때



(b) 유사도가 0.4 일 때



(c) 유사도가 0.6 일 때



(d) 유사도가 0.8 일 때

<그림 5> 유사도, 사전학습기간, 이웃고객의 수에 따른 추천 정확도의 비교

를 설정하고 제시하는 추천시스템을 활용하면 시간이 경과할수록 베스트셀러 기반 추천시스템 보다 더욱 높은 추천 성능을 얻을 수 있을 것으로 판단된다.

6. 결론

본 연구에서는 사회네트워크 분석에서 관계·구조적 특성을 분석하기 위해 널리 활용되고 있는 중심성 개념을 협업필터링에 적용하여 신규고객의 이웃고객을 찾고 그 이웃고객들의 구매정보를 이용하여 신규고객에게 상품을 추천하는 방법을 제시하고, 실험을 통해 그 성능을 검증하였다. 실험 결과, 제시한 추천방법은 기존의 신규고객 추천 방법들과 비교하여 추천의 정확도는 높으면서도, 구매정보 외에 인구통계정보 등과 같은 추가 정보가 필요하지 않으며, 추천 전에 고객이 능동적으로 질의에 응답할 필요가 없는 새로운 방법으로 판명되었다.

본 연구의 한계 및 추후 연구과제로는 다음과 같은 것들이 있다. 첫째, 제시한 추천방법은 같은 시기에 들어온 모든 신규고객에게 같은 상품을 추천하게 된다. 즉, 신규고객에게 개인화된 상품을 추천할 수 없다는 한계점이 있다. 둘째, 이 연구에서는 사회네트워크 분석 기법 중 중심성 개념만을 사용하여 추천방법을 제시하였다. 하지만, 밀도, 집중도 등 사회네트워크의 전반적 성격을 규명하는 분석기법을 활용하여 고객네트워크의 성격을 규정하고 고객네트워크의 성격에 따라 추천방식을 달리하는 연구도 흥미로울 것이다. 셋째, CF의 희박성 문제의 특별한 예가 바로 신규고객 추천문제이다(Huang, et al., 2004). 따라서 본 연구에서 제시하는 방법을 신규고객 뿐만 아니라 구매 기록이 적은 고객들에게도 적용하면 CF의 희박성 문

제까지도 해결할 수 있을 것으로 판단된다. 향후에 실험을 통해 이를 검증하고자 한다. 넷째, 이 연구에서는 과거 구매 데이터를 이용하여 제시한 방법의 추천 정확도를 측정함으로써 그 성능을 평가하였다. 하지만, 실제 고객들은 보다 다양한 상황에서 상품에 대한 구매결정을 내리므로 향후의 연구에서는 실제 고객을 대상으로 한 실험을 통해 제시한 방법을 검증할 필요가 있다.

참고문헌

- 김용학, *사회연결망 분석*, 박영사, 2003.
- 김재경, 조운호, 김승태, 김혜경, “모바일 전자상거래 환경에 적합한 개인화된 추천시스템”, *경영정보학연구*, 15권 3호(2005), 223~241.
- 손동원, *사회 네트워크 분석*, 경문사, 2002.
- 안현철, 한인구, 김경재, “연관규칙기법과 분류모형을 결합한 상품추천 시스템 : G 인터넷 쇼핑몰의 사례”, *경영정보학연구*, 8권 1호(2006), 181~201.
- 윤성준, “데이터마이닝 기법을 통한 백화점의 고객이탈예측 모형 연구”, *한국마케팅저널*, 6권 4호(2005), 45~72.
- 조운호, 박수경, 안도현, 김재경, “재구성된 제품계층도를 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.
- 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 Kno-*

- wledge and Data Engineering*, Vol.17, No. 6(2005), 734~749.
- Aggarwal, C. C., Z. Sun, and P. S. Yu, "Online Algorithms for Finding Profile Association Rules", *Proceedings of the seventh international conference on Information and Knowledge Management*, (1998), 86~95.
- Ahn, H. J., "A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem", *Information Sciences*, Vol.178(2008), 37~51.
- Albert, R. and Barabasi, A. L., "Statistical mechanics 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*, Vol.92(1987), 1170~1182.
- 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.
- Kauffman, S., *The Origins of Order*, Oxford University Press, 1993.
- Kim, D. and B. J. Yum, "Collabrative Filtering based on Iterative Principal Component Analysis", *Expert System with Applications*, Vol.28(2005), 823~830.
- Krulwich, B., "Lifestyle Finder : Intelligent User Profiling Using Large-Scale Demographic Data", *Artificial Intelligence Magazine*, Vol. 18, No.2(1997), 37~45.
- 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'06*, 2006.
- Ryu, Y. U., H. K. Kim, Y. H. Cho, and J. K. Kim, "Peer-oriented content recommendation in a social network", *Proceedings of the Sixteenth Workshop on Information Technologies and Systems*, 2006, 115~120.
- 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'02*, 2002.
- Wasserman, S., *Social network analysis : methods and applications*, Cambridge University Press, 1994.
- Yu, K., A. Schwaighofer, V. Tresp, X. Xu, and H. Kriegel, "Probabilistic Memory-based Collaborative Filtering", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 16, No.1(2004), 56~69.

Abstract

Social Network : A Novel Approach to New Customer Recommendations

Jong Hak Park^{*} · Yoon Ho Cho^{**} · Jae Kyeong Kim^{***}

Collaborative filtering recommends products using customers' preferences, so it cannot recommend products to the new customer who has no preference information. This paper proposes a novel approach to new customer recommendations using the social network analysis which is used to search relationships among social entities such as genetics network, traffic network, organization network, etc. The proposed recommendation method identifies customers most likely to be neighbors to the new customer using the centrality theory in social network analysis and recommends products those customers have liked in the past. The procedure of our method is divided into four phases : purchase similarity analysis, social network construction, centrality-based neighborhood formation, and recommendation generation. To evaluate the effectiveness of our approach, we have conducted several experiments using a data set from a department store in Korea. Our method was compared with the best-seller-based method that uses the best-seller list to generate recommendations for the new customer. The experimental results show that our approach significantly outperforms the best-seller-based method as measured by F1-measure.

Key Words : Social Network, New Customer Recommendation, Cold-Start Recommendation Problem, Centrality Analysis, Collaborative Filtering

* Department of Internet Business, Dongyang Technical College

** School of Business Administration, Kookmin University

*** Management Research Institute and School of Management, KyungHee University

저자 소개



박종학

현재 동양공업전문대학 인터넷비즈니스과 부교수로 재직 중이다. 서울대학교 계산통계학과(전산학전공)에서 학사, KAIST 경영정보공학과에서 석사를 취득하였으며, KAIST 테크노경영대학원에서 박사과정을 수료하였다. LG전자(주)에서 5년간 주임연구원으로 재직한 바 있으며 주 연구 관심분야는 전자상거래, 웹 데이터 마이닝 등이다.



조윤희

현재 국민대학교 경영대학 경영학부 부교수로 재직 중이다. 서울대학교 계산통계학과(전산학전공)를 졸업하고, KAIST 경영정보공학과에서 석사, KAIST 경영공학과에서 박사학위를 취득하였으며, LG전자(주)에서 6년간 주임연구원으로 재직하였다. 주요 연구분야는 Recommender System, Mobile Business, CRM, Data mining 등이며, IEEE Intelligent Systems, Electronic Commerce Research and Applications, Computers and Industrial Engineering, Expert Systems with Applications, International Journal of Internet and Enterprise Management 등에 논문을 게재하였다.



김재경

서울대학교에서 산업공학 학사, 한국과학기술원에서 경영정보시스템 전공으로 석사 및 박사학위를 취득하였으며 현재 경희대학교 경영대학 교수로 재직하고 있다. 미국 텍사스 주립대학교(달라스)에서 교환교수를 역임하였다. 주요 관심분야로는 비즈니스 인텔리전스, 추천시스템, 유비쿼터스 서비스 컴퓨팅 등이다.