

인접성 데이터를 이용한 추천시스템

김진화¹ · 변현수²

¹서강대학교 경영학과 · ²백석예술대학 경영행정학부

접수 2010년 10월 17일, 수정 2010년 11월 25일, 게재확정 2011년 12월 06일

요약

온라인 사용자에게 선택의 어려움을 줄여주고 사용의도를 높이기 위해 만들어진 것이 추천시스템이다. 추천시스템은 정보검색과 정보필터링을 용이하게 하고, 정보 과잉의 문제를 해결하는 데에 많은 도움을 주고 있다. 본 연구의 목적은 웹 상점을 이용하는 사용자들의 클릭스트림 데이터를 분석하여 데이터 인접성의 차이를 확인하고, 이를 통해 상품추천을 제안하고자 하는 데에 있다. 본 연구에서 제안하는 추천시스템의 성과를 검증하기 위하여 실험을 통해 알아본 결과, 추천시스템 적용 전보다 적용 후에 사용자들의 구매 의도는 높아졌고 탐색시간은 줄어들었다.

주요용어: 데이터 인접성, 웹 상점, 추천시스템, 클릭스트림 데이터.

1. 서론

인터넷은 기업이나 일반 소비자의 거래에 있어서 중요한 요소로 작용하고 있고 제품유통 방식에도 큰 변화를 유발시키고 있다. 이에 따라 인터넷 상에서 거래와 주문이 이루어지는 웹 상점 (Web Stores)이 많은 각광을 받게 되었으며, 온라인 유통경로의 중요한 요소로 인정받고 있다. 아직은 웹 상점과 일반 소비자와의 상거래를 의미하는 B2C 전자상거래의 규모가 B2B, B2G 전자상거래 보다 규모가 작다. 그러나 온라인 소매업의 규모가 지속적으로 증가하고 있으며 일반 소비자와 밀접한 관계를 가지고 있음을 상정하면 그 영향력이 더욱 크다고 할 수 있다. 실제로 통계청이 발표한 ‘2010년 2/4분기 전자상거래 및 사이버쇼핑 동향’에 따르면 2010년 2분기 전자상거래 총 거래액은 약 205조원이라고 하였다. 유의할 대목은 2001년 관련 통계조사 시작이후 처음으로 분기별 전자상거래 총 거래액이 200조원을 돌파하였으며, 특히 기업-소비자 간 전자상거래 (B2C)가 전년도와 비교했을 때 34.6%가 증가하여 기업 간 전자상거래 (B2B)와 기업-정부 간 전자상거래 (B2G)에 비해 뚜렷한 상승세를 보이고 있다는 점이다. 이러한 현황을 볼 때 전자상거래 시장, 특히 소매업과 관련된 웹 상점은 발전가능성이 매우 높다고 볼 수 있으며, 웹 상점의 성공을 위해서는 고객으로 하여금 온라인 구매를 많이 활용하도록 다양한 제품 및 서비스, 전략 등을 제공할 필요가 있다.

전자상거래 환경에서는 고객과 상점 간의 상호작용을 기대하기 어렵고 지나치게 많은 상품 및 서비스에 대한 정보가 범람하고 있어서 고객 충성도가 매우 낮은 실정이다. 따라서 보다 현실적인 고객관계 관리리를 위해서 정보의 과부하를 줄이고, 고객과의 일대일 마케팅 전략을 지원하기 위한 정보기술의 지원이 요구된다. 이와 같은 온라인 상거래의 문제점을 극복하고 구매를 촉진시키기 위해서 개발된 것이 추천시스템이다. 추천시스템은 확인하고자 하는 대상과 관련된 정보검색과 정보필터링을 용이하게 하고,

¹ (121-742) 서울특별시 마포구 신수동 1번지, 서강대학교 경영학과, 교수.

² 교신저자: (137-848) 서울특별시 서초구 방배3동 981-7, 백석예술대학 경영행정학부, 전임강사.

E-mail: elbim@hanmail.net

정보 과잉의 문제를 해결하는 데에 많은 도움을 줌으로써, 소비자로서 하여금 웹 상점의 이용과 구매를 높이도록 이끄는 수단이 된다. 예를 들면 생명보험사의 텔레마케팅 실행 시에 적용하거나 (고봉성 등, 2009), 쇼핑몰의 이메일 캠페인 진행에 활용해본 결과 (김연형과 이석원, 2009) 그 필요성이 인정되고 있다. 현실적으로 기존의 여러 추천시스템들도 나름대로의 성과를 거두고 있으며 Amazon.com과 같은 유수의 전자상거래 업체에서도 적극적으로 반영되어 기업수익에 긍정적인 효과를 주고 있다.

기존의 추천시스템은 그 성과에도 불구하고 몇 가지 문제점이 제시되고 있다. 그 중 하나는 연관성을 기반으로 추천하는 경우 희박성 (sparsity)의 문제가 발생한다는 점이다. 구매를 상대적으로 적게 하는 고객이나 구매가 상대적으로 적게 이루어지는 상품에 관해서는 제대로 된 추천을 하지 못할 가능성이 있다. 다른 하나는 협업 필터링을 기반으로 하는 추천의 경우 확장성 (scalability)의 문제가 발생한다. 거래가 증가할수록 거래 데이터가 증가함에 따라 유사성 계산에 따른 연산량이 기하급수적으로 증가하기 때문이다 (김경재와 김병국, 2005). 본 연구에서 제안하는 상품추천시스템은 사용자가 웹 상점에서 활동한 데이터 이력을 이용하여 그 활동데이터 간의 인접성 차이를 확인한다. 전통적인 오프라인 상점에서의 구매데이터는 구매행동에 관한 정보만을 얻을 수 있지만, 사용자의 활동내역이 담긴 클릭스트림 데이터는 비 구매 소비자의 정보에 대해서도 제공해 주기 때문이다 (Moe와 Fader, 2004). 클릭스트림 데이터를 통해서 소비자의 여러 가지 행동유형을 분석하는 것은 마케팅 전략이나 경쟁전략을 수립하는데 유용한 방법이 될 수 있다.

본 연구에서는 사용자의 클릭스트림 데이터를 분석하여 그래프 이론에 기반을 둔 인접행렬을 작성하고, 여기서 추출한 인접성 차이의 변동을 확인함으로써 사용자에게 적합한 상품을 추천하는 방법을 제안하고자 한다. 또한 본 연구의 다른 특징으로는 클릭스트림 데이터를 이용하기 때문에 분석 데이터 자체가 순차적 성격을 띠고 있으며, 사용자로서 하여금 사용자의 프로파일이나 특정 선호도 등과 같은 추가적인 자료입력 절차를 요구하지 않는 장점도 있다.

2. 관련 연구

2.1. 추천시스템

사용자 개인의 취향과 기호에 맞추어진 상품이나 서비스를 제공하는 개인화 개념이 구체적으로 실현된 것 중에 추천시스템이 있다. 정보처리 부담을 덜고 고객과의 관계를 강화하기 위한 추천시스템은 통계적 기법과 지식탐사 기술의 도움을 받아 많이 사용되고 있으며 구매촉진을 지원하기 위한 방법으로도 이용된다 (김재경 등, 2005).

추천시스템을 위한 데이터 분석방법으로써 Eirinaki와 Vazirgiannis (2003)는 콘텐츠 기반 필터링, 협업 필터링, 규칙 기반 필터링, 그리고 웹 사용 마이닝 등 크게 4가지로 분류하였다. 콘텐츠 기반 필터링은 개별 사용자의 선호도에만 초점을 맞추어서 사용자의 과거 행동패턴에 맞는 항목을 추천하는 시스템이다. 협업 필터링은 비슷한 행동패턴을 가지는 사용자들은 결국 관심이 유사할 것이고 이에 따라 어떤 항목에 유사한 평가를 내릴 것으로 예측하는 데에서 출발한다. 따라서 어떤 상품에 대한 평가를 비슷하게 내린 집단, 즉 선호도나 관심이 거의 동등한 집단에게 상품정보를 제공하는 시스템이다. 규칙 기반 필터링에서는 사용자들에게 몇 가지의 질문을 주고 이를 토대로 의사결정나무를 추출한다. 이를 통해 사용자 개개인의 필요에 적합한 상품을 결과로 제공하는 시스템이다. 끝으로 웹 사용 마이닝은 웹 로그 데이터에 대한 통계적 또는 데이터 마이닝적 방법을 적용하여 사용자의 행동에 대한 유용한 패턴을 알아내려는 분석기법이다 (Eirinaki와 Vazirgiannis, 2003).

현재의 연구흐름을 살펴보면 초기 콘텐츠 기반의 필터링을 많이 발전 및 보완시킨 협업 필터링의 개념을 많이 활용하고 있다. 협업 필터링에서는 어떻게 이웃고객을 규정하고 그 수를 조절할 것인지가 주요 관건이 되는데, 선호도 예측을 위한 보정함수를 개발하거나 (이희춘, 2009), 적정한 이웃의 수를 규정하

는 예측식을 제안하는 방법 (이석준, 2009) 등이 보고되어 있다. 추천시스템들은 각자의 장단점이 있기 때문에 효용을 극대화하기 위하여 혼합형을 취하기도 한다. 예를 들자면 협업필터링과 콘텐츠 기반 필터링을 결합한 Fab 시스템 등이다 (Balabanovic과 Shoham, 1997).

2.2. 클릭스트림 데이터 분석

온라인 사용자들의 활동내역을 기록한 데이터를 클릭스트림 데이터 (clickstream data)라고 하며 이에 대한 분석 및 활용방법이 많이 제시되고 있다. 클릭스트림 데이터는 전자상거래 사이트 등 온라인 웹 사이트에서의 사용자 행동을 담은 이력을 추적하고 저장함으로써 생성되는 상세한 전자적 정보를 의미한다 (Bucklin 등, 2002). 이러한 클릭스트림 데이터에 대한 연구 분야는 개인의 목적 및 선택결정 영역에 따라 다음의 4가지로 세분화할 수 있다 (Bucklin 등, 2002).

표 2.1 인터넷 선택방식에 따른 연구방법

개인의 목적	선택결정 영역	
	웹사이트 내부	웹사이트 내·외부
검색	사이트 향해	사이트 선택
구매	전자상거래 또는 추천시스템	ShopBots

표 2.1에서 개인의 목적은 사용자의 웹사이트 방문 목적이 검색 (또는 탐색), 구매 중 어떠한 것인지를 파악하는 것이다. 그리고 선택결정 영역은 단일 웹사이트 내부에서 의사결정이 이루어지는지 혹은 여러 개의 웹사이트를 돌아다닌 후에 이루어지는지를 확인하는 것이다. 표 2.1에서 ShopBots (shorthand for shopping robots)라는 것은 사용자의 의사결정을 도와주는 일종의 에이전트 (agent)로서 제품 및 가격정보를 대안으로 제공하는 역할을 한다.

온라인 쇼핑과 오프라인 쇼핑의 큰 차이점으로 이동비용의 차이를 들 수 있다. 오프라인 상점에서는 상점 그 자체 또는 제품의 선택에 있어서 상점으로서의 이동비용이 큰 영향을 미치지만, 시공간의 한계가 거의 없는 온라인 상점의 경우에는 이동비용이 적거나 거의 없다. 이 때문에 사용자는 온라인 쇼핑 시 구매결정을 나중에 내리거나 재방문하는 경향이 높다. 그러한 온라인 사용자의 성향 때문에 온라인 상점에서는 실제 구매와 관련이 적은 비 구매 자료가 오프라인 상점보다도 활용도가 높다 (Moe와 Fader, 2004). 이처럼 클릭스트림 데이터에는 일상적인 경로로 얻기 힘든 사용자의 행동방식이 많이 숨겨져 있기 때문에 발전된 분석기법 및 예측방법이 제시된다면 더 높은 수익이 전자상거래 사이트에 제공될 수 있다 (Moe와 Fader, 2002).

클릭스트림 데이터는 다음의 방법들을 통해 수집될 수 있다. 첫째, 온라인 벤더들의 웹서버에 저장된 로그파일, 즉 소매업자가 보유한 데이터를 획득한다. 둘째, 사용자의 행동을 돕는 ShopBots 에이전트가 제공해 주는 데이터를 이용한다. 셋째, 실험실에서 실험디자인을 통해 인위적으로 데이터를 수집 및 분석한다. 넷째, 사용자 패널 데이터로서 컴퓨터에 소프트웨어를 설치해서 사용자들의 활동내역을 수집하는 방법이 있는데 시장조사업체 등에서 널리 사용되고 있다 (신일순 등, 2002).

2.3. 데이터 인접성

데이터의 인접성 (adjacency)은 의사결정 공간 (decision space)에서 유용하게 쓰일 수 있다. 어떤 품목 i 와 품목 j 가 동시에 구매되었는지, 또는 품목 i 의 구입이 품목 j 의 구매에 영향을 주었는지의 여부가 데이터의 인접성이라고 할 수 있다. 데이터 인접성을 토대로 연관규칙 (association rules) 분석, 추천 프로그램 (recommendation applications) 개발, 시각화 모형 (visualization model) 구축 등에 적용하고 있다.

인터넷은 여러 통신망의 집합체이며, 웹에 퍼져 있는 수많은 개체들이 서로 연결되어 있다. 인터넷 사용자들은 어떤 순서와 방향을 가지면서 활동한다. 따라서 인터넷 공간에 인접행렬의 개념을 이용할 수 있다 (Błazewicz 등, 2005). 어떤 인접행렬 A 는 n 개의 점을 가지는 $n \times n$ 정방행렬이다. 인접행렬의 원소 $a_{ij} = 1$ 이면 두 점이 인접해 있다는 것이며, $a_{ij} = 0$ 이면 두 점은 인접해 있지 않은 것이다.

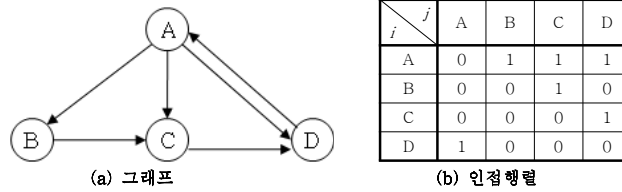


그림 2.1 그래프와 그 그래프를 토대로 작성한 인접행렬의 예

이러한 인접행렬에 기반을 두고 웹사이트를 구성하는 여러 단위들을 하나의 점으로 놓은 다음 이에 방향성을 부여한 뒤, 각 사용자의 방문행적을 확인할 수 있다. 기본적인 그래프와 인접행렬의 예는 그림 2.1과 같다. 그림 2.1의 (a)를 보면 유향성 (directness)을 지니고 있음을 나타내기 위해 화살표로 방향성을 표시한 그래프의 예가 나타나 있다. 또한 그림 2.1의 (b)에는 그래프 (a)를 토대로 작성한 인접행렬이 나타나 있다. 인접행렬은 기본적으로 0과 1의 값을 가지며 횡수를 부여할 수도 있다. 방향성을 인식하기 위해 행에서 열로 이동한다고 규정하기로 한다 ($i \rightarrow j$).

2.4. 다차원 척도분석

다차원 척도분석 (MDS: multidimensional scaling)은 n 개의 대상이 있을 때, 그 대상에 대한 변수를 측정하고 그 변수를 이용하여 대상물 사이의 거리 또는 (비)유사성을 측정하여 대상물들의 공간적 구조를 설명하는 통계적 방법이다. 다차원 척도분석은 정신분석학에 그 기원을 두고 있으며, 데이터의 구성에 따라 2차원이나 3차원의 공간구조를 통해서도 확인해 볼 수 있다 (Shang 등, 2004).

가장 기본적인 거리계산은 2차원 좌표상에 있는 두 점 사이의 거리인 유클리드 거리계산을 거론할 수 있다. 어떤 2차원 좌표위에 두 점 X_i 와 X_j 가 있다고 가정한다. 그러면 $X_i = (x_{i1}, x_{i2}, \dots, x_{im})$ 와 $X_j = (x_{j1}, x_{j2}, \dots, x_{jm})$ 의 유클리드 거리를 \widehat{d}_{ij} 라 하면 이것을 식 (2.1)과 같이 표현할 수 있다.

$$\widehat{d}_{ij} = \sqrt{\sum_{k=1}^m (x_{ik} - x_{jk})^2} \quad (2.1)$$

여기서 m 은 차원을 의미하는데 통상적으로 시각화를 위해서 2 또는 3을 채택하는 경우가 많다. 이와 같이 다차원 척도분석에서는 관측대상 간의 유사성 또는 비유사성을 판별하고 그것을 측정하는 것이 중요하다. 본 연구에서는 어떤 인접행렬 D 가 있을 때 품목 i 와 품목 j 의 유사성은 행렬의 원소 값 a_{ij} 로 규정할 수 있다. 다차원 척도분석에서 대상물 사이의 유사성 혹은 비유사성 측정 시 정확도를 높이기 위해서 스트레스 값을 통해 적합성을 판별하곤 한다. 스트레스 값은 실제 거리와 추정 거리 사이의 오차

를 나타내며 작으면 작을수록 좋다. 일반적인 스트레스 함수 값은 식 (2.2)와 같이 구한다.

$$S = \sqrt{\frac{\sum_{1=i, i \neq j}^n (d_{ij} - \widehat{d}_{ij})^2}{\sum_{1=i, i \neq j}^n (d_{ij})^2}} \tag{2.2}$$

d_{ij} = 관측대상물인 i부터 j까지의 거리

\widehat{d}_{ij} = 프로그램이 추정한 거리

다차원 척도분석에서는 반복계산을 통해 가장 적합한 공간구조를 분석해주며 이러한 반복횟수나 결과 값은 초기수치의 품질에 영향을 많이 받는 편이다.

3. 실험설계 및 실증분석

3.1. 실험설계 과정

본 연구를 위한 데이터를 수집하기 위하여 가상 웹 상점을 구축하였으며, 대학의 학부과정 학생 및 대학원생을 피 실험자로 모집하여 해당 웹 상점에 접속하여 활동하도록 하였다. 이들의 활동기록은 구축한 가상 웹 상점의 서버에 저장되었으며, 실험대상자들도 자신의 활동이력을 확인할 수 있도록 연습을 거쳤다.

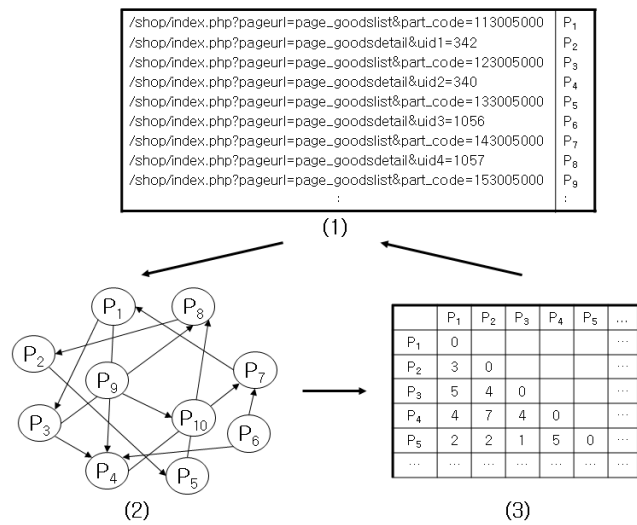


그림 3.1 연구단계 서술

조사기간 동안 획득한 클릭스트림 데이터를 대상으로 인접행렬을 작성해 보았다. 클릭스트림 데이터 자체는 길고 알아보기 어려우므로, 인접행렬을 만들기 위해서 클릭스트림 데이터 중 제품추천에 필요한 상품에 해당하는 부분을 P_n 의 형태로 변환시켰다. 이 P_n 은 연결그래프에서 하나의 점이 된다. 웹과 웹 사이에서 뿐만 아니라 어떤 웹사이트 내에서 사용자의 이동은 자유롭고 활발하다는 점에서 유향성을 지닌 연결그래프가 적당하다고 할 수 있다. 클릭스트림 데이터 각각에 부여된 P_n 에 따라 얻어진 정보

를 토대로 해서 구현한 인접행렬은 상품갯수 n 에 따라 결정되는 $n \times n$ 정방행렬이며 유향그래프를 토대로 작성된 행렬이고, 행에서 열 방향으로 이동한다고 규정해 둔다.

그림 3.1에서 제시하는 본 연구의 단계별 과정은 다음과 같다. (1) 실험대상자들로 하여금 본 연구를 위해 구축한 웹 쇼핑몰에 접속하여 활동하게 한다. 그러면 피 실험자들의 활동내역인 클릭스트림 데이터를 수집할 수 있으며 이때 각각의 상품을 P_n 과 같이 변환하여 이후의 분석을 용이하게 한다. (2) 사용자의 활동이력을 그래프로 치환하여 제품과 제품 사이의 연결정도를 확인한다. 어떤 사용자의 웹사이트 내 활동은 순차적이면서 자유로운 흐름을 가지고 있기 때문에 구축되는 그래프의 모양 또한 다소 복잡하게 나타나게 된다. (3) 사용자들이 방향성을 가지고 움직인다고 가정하고 있기 때문에 인접행렬은 비대칭행렬로 나타나게 되어 $P_{ij} \neq P_{ji}$ 와 같이 된다. 이러한 경우 두 제품 P_i 와 P_j 의 유사성을 구하는 방식으로 Condon 등 (2002)은 두 원소 값 P_{ij} 와 P_{ji} 의 평균을 취하고 있으며, 본 연구에서도 두 원소 값의 평균으로 두 점 사이의 유사성을 측정하기로 한다.

Derived Stimulus Configuration

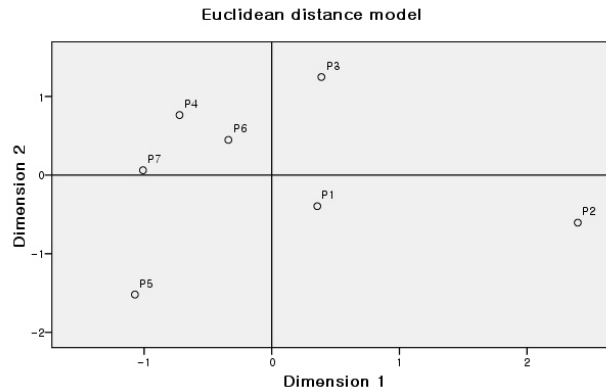


그림 3.2 본 연구에서의 다차원 척도분석 결과 사례

이러한 과정을 통해 구축한 인접행렬을 이용하여 각 품목 간 유사성을 확인한 다음 이것을 토대로 다차원 척도분석을 실행하였다. 다차원 척도분석 결과 각각의 상품들끼리의 유사성이 다차원 공간에 구축되어 시각적으로 확인할 수 있다. 이를 토대로 어떤 상품과 가까운 유사성을 가지고 있는 제품을 사용자에게 추천하는 것이다. 다차원 척도분석 결과 중 하나의 예가 그림 3.2에 제시되어 있다. 그림 3.2에서는 각 제품 간 유사성의 범위를 알 수 있는 하나의 예를 보이고 있다. 그림 3.2의 경우 다른 제품에 비해 P_4, P_6, P_7 이 특히 가까운 거리에서 군집을 이루고 있다. 따라서 어떤 사용자가 (P_4, P_6, P_7) 중 하나의 제품을 선택하면 나머지 제품을 같이 소개하는 형태로 추천이 이루어진다.

인접행렬의 값을 토대로 해서 실행한 다차원 척도분석의 모형 적합도 $RSQ = 0.64608$ 이 제시되었는데 일반적으로 RSQ 가 0.6이상이면 유의하다고 알려져 있다. 따라서 인접성 데이터를 이용한 다차원분석은 통계적으로는 유의한 수준으로 인식할 수 있다. 또한 스트레스 값은 Kruskal의 기준을 많이 따르는 데 스트레스 값이 0.05이하이면 적합도가 높고, 0.20이상이면 적합도가 나쁘다고 할 수 있다 (Kruskal과 Wish, 1991). 본 연구에서의 스트레스 값은 0.15858로 받아들일 수 있는 수준으로 드러났다. 모형 적합도 RSQ 와 스트레스 값의 수치가 크게 만족스러운 수준이 아닌 것은 우선 그 대상이 되는 품목이 너무 많고, 연관성 있는 두 품목 사이에 실험대상자의 순차적 이동으로 인한 중간과정이 개입되

어 있기 때문에 판단된다.

3.2. 유용성 검증

본 연구에서 제안하는 추천시스템의 유용성을 검증하기 위하여 그 효과를 분석하였다. 실증분석은 다음과 같은 순서로 진행되었다. 처음 클릭스트림 데이터 수집을 위하여 구축한 가상의 웹 상점에서 얻은 자료를 토대로 구축한 추천시스템을 평가받기 위해 새롭게 설문대상자들을 모집하였다. 설문에 동의한 100명의 새로운 피 실험자들로 하여금 본 연구를 위해 구축한 웹 상점을 방문토록 하였다. 피 실험자들의 활동내역을 기반으로 인접행렬을 만들고 인접성 차이의 변동을 알아보았다. 그 결과를 토대로 연구 대상 웹 상점에서 제시하는 상품들을 추천하도록 웹사이트를 구현하였다. 여기까지의 과정에서 실험의 진행에 따라오지 못하는 피 실험자와 지나친 이상 값을 보이는 피 실험자를 제외하고 총 88명을 대상으로 실험을 계속 진행하였다. 이 88명의 피 실험자를 대상으로 연구자가 제안한 기법의 상품추천을 지원하는 웹 상점을 방문하는 경우 추천방식의 유용성을 확인하고자 하였다.

표 3.1 추천시스템 적용 전과 적용 후의 paired-t test 결과

		평균	참가자 수	t값	p값
구매의도 (5점 척도)	추천 전	2.23	88	2.94	0.000
	추천 후	3.05	88		
탐색시간 (초)	추천 전	607	88	4.01	0.000
	추천 후	545	88		

본 연구에서 제안하는 상품추천 방식의 적용 전과 적용 후에 각각 구매의도와 탐색시간을 측정하여 상품추천 전과 상품추천 후의 변화를 비교하였다. 이 때 피 실험자에게 발생할 수 있는 학습효과를 배제하기 위해 사전에 본 연구에서 제시하는 추천시스템에 대해서는 언급하거나 보여주지 않았고, 추천 전과 추천 후에 1시간 이상의 간격을 둬으로써 학습효과를 제거하기 위한 절차를 삽입하였다. 상품추천 적용 전과 후에 구매의도와 탐색시간이 유의하게 변화하였는지 알아보기 위하여, 동일 집단의 평균값 차이를 알 수 있는 paired-t 검정을 실시하였다. 그 실험결과가 표 3.1에 나타나 있다.

분석결과 피 실험자 집단은 추천시스템 적용 후에 구매의도가 증가하였으며 탐색시간이 줄어들었다. 그리고 그러한 변화의 차이는 유의수준 $\alpha = 0.05$ 에서 모두 충분히 유의함을 알 수 있다. 따라서 본 연구에서 제시하는 추천시스템의 적용으로 웹 상점 탐색시 보다 빠르고 편리한 웹사이트 이용에 도움을 주고 있다고 할 수 있다.

4. 결론 및 향후 과제

인터넷 상에서 거래가 이루어지는 전자상거래의 발전가능성은 무궁무진하며 실제로 많은 성과를 보이고 있다. 이러한 전자상거래 시장의 최일선에 존재하며 실제 상거래가 발생하는 곳이 웹 상점이다. 특히 일반 소비자와의 접점이 되는 기업-소비자간 전자상거래 (B2C)에 있어서 웹 상점은 가장 관심과 초점의 대상이 되지만, 아직 웹 상점을 이용하는 사용자들의 구매 전환율이나 충성도는 매우 낮은 편이다. 웹 상점의 발전이 더딘 이유로는 웹사이트 자체의 구조적 문제, 적은 탐색비용, 쉬운 웹사이트 비교, 부족한 제품 및 서비스 등이 거론된다. 이러한 문제점들은 여러 연구와 방법론에 의해 해소되고 있으며, 웹 상점에 대한 양질의 발전이 앞으로 기대되고 있다.

사용자의 구매활동을 높이기 위해서 제시된 많은 방법들이 있고, 이들 중에서 추천시스템은 정보 과잉을 줄이고 정보검색과 정보필터링에 유용한 면이 있다. 본 연구에서는 인접성 데이터를 이용한 추천시스템을 고안하고 그 방법을 제안하였다. 이를 위해 실험대상자들을 선별하여 실험설계 방법으로 웹 상

점을 구축하였고, 이 웹 상점을 방문하는 사용자들에 대한 클릭스트림 데이터를 획득하였다. 그 데이터 셋을 기반으로 하여 품목 간 연관성을 보여주는 그래프와 인접행렬을 구축하였고 이를 바탕으로 각 제품 간 인접성의 차이와 그 변동을 알아보았다. 인터넷은 많은 개체의 집합으로 이루어져 있기 때문에 데이터 인접성의 개념을 적용하기에 알맞기 때문이다. 또한 실증연구를 통하여 본 추천시스템의 효과를 분석하였다.

본 연구에서 제안하는 추천시스템의 유용성을 검증하기 위하여 실증연구를 실행하였다. 연구결과 추천시스템 적용 전보다 적용 후에 구매 의도는 높아졌고 탐색시간은 줄어들었다. 따라서 본 연구에서 제안하는 추천시스템의 효용성이 있다고 할 수 있다. 향후 연구에서는 무엇보다도 클릭스트림 데이터의 수집과 정제에 보다 면밀한 주의를 기울일 필요가 있다. 또한 그래프 이론을 통한 인접행렬 구축 시 서로 다른 그래프의 종류를 반영해 볼 수도 있다. 그리고 본 연구에서는 추천시스템의 적용전과 적용후의 구매의도 및 탐색시간 변화를 통해 시스템 유용성을 검증했는데 이때 발생할 수 있는 학습효과를 고려해서 대처할 수 있는 더 나은 방법을 고안할 필요가 있다. 이러한 점들을 개선한 뒤 추가분석이 이루어지면 보다 좋은 결과가 도출될 것이다.

참고문헌

- 고봉성, 이석원, 허정 (2009). 생명보험사 텔레마케팅 효율성 제고에 관한연구. <한국데이터정보과학회지>, **20**, 673-684.
- 김경재, 김병국 (2005). 데이터 마이닝을 이용한 인터넷 소규모 상품추천시스템. <한국지능정보시스템학회논문지>, **11**, 191-205.
- 김연형, 이석원 (2009). 추천시스템을 이용한 이메일 효율성 제고에 관한 연구. <한국데이터정보과학회지>, **20**, 1129-1143.
- 김재경, 안도현, 조윤호 (2005). 개인별 상품추천시스템, WebCF-PT: 웹마이닝과 상품계층도를 이용한 협업필터링. <경영정보학연구>, **15**, 63-79.
- 신일순, 정부연, 김보은 (2002). <패널데이터를 이용한 e-Business 소비자행태 분석>, 연구보고 02-15, 정보통신정책연구원
- 이석준 (2009). 근접 이웃 선정 협력적 필터링 추천시스템에서 이웃 선정 방법에 관한 연구. <한국데이터정보과학회지>, **20**, 809-818.
- 이희춘 (2009). 협력적 필터링 추천기법에서 이웃 수를 이용한 선호도 예측 정확도 향상. <한국데이터정보과학회지>, **20**, 505-514.
- Balabanovic, M. and Shoham, Y. (1997). Fab: Content-based, collaborative recommendation. *Communications of the ACM*, **40**, 66-72.
- Błazewicz, J., Pesch, E. and Sterna, M. (2005). A novel representation of graph structures in web mining and data analysis. *Omega*, **33**, 65-71.
- Bucklin, R. E., Lattin, J. M., Ansari, A., Gupta, S., Bell, D., Coupey, E., Little, J. D. C., Mela, C., Montgomery, A. and Steckel, J. (2002). Choice and the internet: From clickstream to research stream. *Marketing Letters*, **13**, 245-258.
- Condon, E., Golden, B., Lele, S., Raghavan, S. and Wasil, E. (2002). A visualization model based on adjacency data. *Decision Support Systems*, **33**, 349-362.
- Eirinaki, M. and Vazirgiannis, M. (2003). Web mining for web personalization. *ACM Transactions on Internet Technology*, **3**, 1-27.
- Kruskal, J. B. and Wish, M. (1991). Multidimensional scaling. Beverly Hills, CA: Sage.
- Moe, W. and Fader, P. S. (2002). Uncovering patterns in cybershopping. *California Management Review*, **43**, 106-117.
- Moe, W. and Fader, P. S. (2004). Dynamic conversion behavior at e-commerce sites. *Management Science*, **50**, 326-335.
- Shang, Y., Ruml, W., Zhang, Y. and Fromherz, M. (2004). Localization from connectivity in sensor networks. *IEEE Transactions on Parallel and Distributed Systems*, **15**, 961-974.

A product recommendation system based on adjacency data

Jinhwa Kim¹ · Hyeonsu Byeon²

¹Sogang Business School, Sogang University

²Department of Management and Administration, Baekseok Arts University

Received 17 October 2010, revised 25 November 2010, accepted 06 December 2011

Abstract

Recommendation systems are developed to overcome the problems of selection and to promote intention to use. In this study, we propose a recommendation system using adjacency data according to user's behavior over time. For this, the product adjacencies are identified from the adjacency matrix based on graph theory. This research finds that there is a trend in the users' behavior over time though product adjacency fluctuates over time. The system is tested on its usability. The tests show that implementing this recommendation system increases users' intention to purchase and reduces the search time.

Keywords: Clickstream data, data adjacency, recommendation system, web store.

¹ Professor, Department of Business Administration, Sogang University, Seoul 121-742, Korea.

² Corresponding author: Full-time Instructor, Department of Management and Administration, Baekseok Arts University, Seoul 137-848, Korea. E-mail: elbim@hanmail.net