

연관 규칙을 활용한 고객구매 제품 분석

조병석

고려대학교 컴퓨터정보통신대학원 소프트웨어공학과

e-mail : gzone@korea.ac.kr

Customer Purchase Behavior Modeling using Association Rule Mining

Cho Byong Sok

Graduate School of Computer and Information Technology, Korea University

요 약

패션 시장은 항상 경쟁이 치열하고 고객의 변화 및 이탈이 심한 시장이다. 경쟁의 요소가 품질 등의 가격적인 요소에서 디자인 및 서비스 등 비 가격 적인 요소의 중요성이 부각되고 있다. 이에 따라 고객 정보에 대한 분석을 기반으로 한 마케팅 및 판매 전략이 중요한 것은 두말할 필요가 없다. 정보 기술과 다양한 분석 기법은 다양한 방법으로 고객의 행동을 분석하여 고객의 구매 형태를 분석 및 예측하여 고객별로 차별화된 마케팅과 서비스를 제공할 수 있도록 한다.

1. 서론

기업이 고객과 상호 교류하는 방식은 최근 몇 년사 이 엄청나게 빠르고 다양하게 변하고 있다. 기업들은 고객들을 더 잘 이해하고 고객들의 Needs 에 대해 더 빨리 대응하는 것이 시장에서 성공의 요인임을 느끼고 있다. 즉, 정보화시대에 기업경영은 전통적인 자원을 효과적으로 운영하는 것 외에도 고객만족을 위해서 제품이나 서비스의 질을 꾸준히 향상시켜야만 하는 것을 두말할 필요가 없다. 이러한 비즈니스 환경을 아래 고객의 요구가 무엇인지를 지속적으로 파악하고 대응하기 위해서 고객정보를 축적하고 분석해야 한다.

이를 바탕으로 CRM 에서는 고객관리 전략을 개발하고 지속적으로 수행하는 능력이 중요하며, 고객 정보를 분석하는 도구로서 데이터 마이닝의 사용이 요구되고 있다.

현재 패션시장은 성장이 둔화되고 있고 경쟁이 치열하여, 그로 인해서 각 회사들은 각 고객군별 차별화된 서비스를 제공하기 위해서 다양한 마케팅 활동을 기획하고 수행한다. 다양한 고객 Segmentation 을 기반으로 각 회사들은 Churn Management, Cross Selling, Up-Selling 의 기존 의 전통적인 CRM 에서 고객의 경험까지 관리하는 CEM(Customer Experience Management)까지 다양한 노력을 수행하고 있다.

2. 연구방향

기업의 입장에서 고객정보를 전략적 자산으로 간주하고 분석하고 이에 대한 마케팅을 수행하는 것은 기

업의 중요한 성공 요소이다. 시장 및 고객의 Needs 가 세분화 되어 짐에 따라 마케팅 및 고객 전략도 세분화 하는 것은 중요한 요소이다. 고객의 생애가치 (LTV-Life Time Value)를 평가하여 그중에서 더 가치 있는 고객에 대해서 차별화된 서비스 및 캠페인을 수행하는 것은 기업의 입장에서도 비용을 줄이고 우수 고객에 더 잘 접근할 수 있게 해야 한다.

먼저 기존의 고객군을 전통적인 RFM (Recency, Frequency, Monetary)를 활용하여 고객을 Grouping 하였고 이에 바탕으로 본 논문에서는 국내 한 패션회사의 고객의 구매 제품간의 연관성을 Association Rule 을 활용하여 분석하고 이를 바탕으로 고객 전략에 대한 방향을 제시하고자 한다.

특정의 제품에서 특정 아이템의 거래가 발생하면 다른 특정 아이템집합의 특정아이템의 거래가 발생하는 현상" 연관성(association) 이라고 하고 이러한 현상을 발견하고자하는 분석을 데이터 마이닝(data mining)기법에서 연관성 분석(association analysis) 또는 연관성규칙 발견 분석(association rule analysis)이라고 한다. [Agrawal et al., 1993; Srikant et al.,1993]

현재 C 사의 전체 브랜드 중 현재 관리 중이며 향후 지속 관리 예정인 브랜드 21 개중 많이 알려진 7 개 브랜드에 대해서 연관분석 (장바구니분석) 을 활용해 분석하고자 한다. 이를 통해서 신규 고객 확보 시나 구매 고객 대상에게 판매 시 소비자의 구매 패턴의 변화를 파악하여 구매 가능성이 높은 연관 제품을 Cross Selling 또는 Up Selling 등의 판매 및 서비스 전략에 활용하여 고객 구매력을 극대화 할 수 있도록

한다

C 사 브랜드는 크게 백화점과 가두점을 중심으로 판매가 발생하는데 본 연구에서는 매출이 더 많은 백화점 고객을 대상으로 한다.

3. 실험 및 분석

B 대표 브랜드(B.Com)는 7 개의 자 브랜드를 가지고 있고 연관성 분석은 다음과 같이 수행하였다.

1. 분석대상 선정
 - A. 분석데이터 기간(03.10~05.09) 2 년
 - B. 두개 이상 브랜드 이용 고객으로 한정
2. 유통매장별 분류
 - A. 가두점(복합매장) 과 백화점(단일매장)
 - B. 본 연구는 백화점 고객만으로 한정
3. 연관 Rule 1 차도출
 - A. 연관성 분석
 - B. 시차 분석(A 구매 후 B 구매)
4. 연관 Rule 최종 도출
 - A. Confidence, Support, Lift 분석

Association Rule 은 아래와 같이 정의된다.

Confidence 는 B 구매 고객 중 A 를 구매하는 확률로 순수한 연관 정도를 파악한다.

$$\text{신뢰도} = \frac{\text{품목 A와 B를 동시에 포함하는 거래 수}}{\text{품목 A를 포함하는 거래 수}}$$

Support 는 전체 고객중 B 와 A 를 구매하는 확률을 의미하고 다음과 같이 구한다.

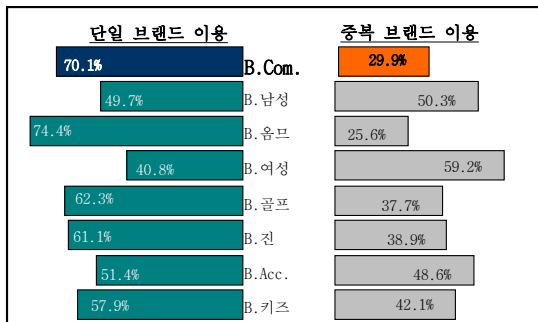
$$\text{지지도} = \frac{\text{품목 A와 B를 동시에 포함하는 거래 수}}{\text{전체 거래 수}}$$

Lift 는 연관브랜드간의 상호 개연성 정도를 파악한다. Lift 가 1 이면 두 브랜드간 연관성이 없음을 의미하고 1 이상이면 긍정적인 연관성을 의미한다.

$$\text{향상도} = \frac{\text{품목 A와 B를 동시에 포함하는 거래 수}}{\text{품목 A 포함 거래 수} * \text{품목 B 포함 거래 수}}$$

분석 결과 B.Com의 7 개 브랜드간의 연관분석을 하기 위해 2 개 이상 브랜드를 이용하는 중복브랜드 현황과 브랜드간 중복현황을 살펴본 결과 전체적으로 단일브랜드 이용 비율이 높지만 중복 브랜드를 이용하는 고객의 비중도 상당하므로 유의한 브랜드간 연관분석 결과를 도출 했다.

[그림 1] B.Com 브랜드 구성 및 중복이용 현황



B.Com 은 그림[1]에서와 같이 7 개의 브랜드로 구성된 다.

백화점 내에 전개된 B.Com 의 연관성 분석을 수행한 결과 연관 분석의 경우 전체적으로 B.남성과 B.Acc, 그리고 B.여성과 B.킴즈/B.진/B.골프 사이의 연관관계가 높으며 시차분석의 결과 B.남성과 B.여성 그리고 B.Acc.사이에서 시차를 이용해 연관 구매하는 고객의 비율이 높은 것으로 나타났다.

[표 1] 두개 브랜드간 연관성 Rule

□ : Confidence □ : Support

	B 남성	B 여성	B 골프	B 진	B.Acc.	B 킴즈	B 옴므
B 남성		15,707	3,045	4,141	68,392	3,194	1,355
B 여성	19.9	14.0	3.9	2.7	5.3	3.7	86.9
B 골프	48.6	2.7	43.4	2.4	10.0	0.6	39.2
B 진	43.2	3.7	44.4	3.8	6.5	0.6	50.4
B A C c.	71.7	61.1	36.3	30.9	2.6	2.2	5.1
B 킴즈	43.2	2.9	47.6	3.1	13.0	0.9	9.9
B 옴므	64.7	1.2	17.2	0.3	11.2	0.2	14.0

백화점에서 판매된 B.Com 을 이용하는 고객의 브랜드간 연관성 분석을 수행한 결과 주로 B.남성, B.Acc, B.킴즈, B.골프, B.진사이의 연관 Rule 이 의미가 있으며 이중 B.남성과 B.Acc.간의 연관구매가 가장 많이 일어나고 있어 연관구매 가능성이 높은 브랜드에 대한 판매 및 마케팅 전략 수립에 활용할 수 있다.

연관 분석 결과는 다음과 같다

[표 2] 2 개 브랜드 간 상세 연관 Rule

Rule	고객수	Confidence	Support	Lift
BP + BE	68392	86.9	61.1	1.0
BE + BP	68392	71.7	61.1	1.0
BG&BE + BL	1238	50.4	1.1	1.3
BK&BE + BL	1748	49.7	1.6	1.2
BK + BL	3522	47.6	3.1	1.2
BJ + BL	4249	44.4	3.8	1.1
BG + BL	2717	43.4	2.4	1.1
BJ & BE + BL	2007	41.6	1.8	1.0
BG + BK	962	15.4	0.9	2.3

분석결과 전체적으로 B.남성과 B.Acc 를 구매하는 고객의 비율이 매우 높고 B.골프와 B.Acc, B.키즈와 B.Acc, B.진과 B.Acc 를 구매하는 고객은 B.여성 브랜드를 구매하는 가능성이 높은 것을 알 수 있다. B.여성은 B.키즈/B.진/B.골프의류와의 구매비율이 높고 B.골프와 B.키즈의 연관구매 비율이 높음을 알 수 있다.

다음은 백화점 전개된 B.Com.의 구매시차에 따른 브랜드간 연관분석을 살펴본 결과 전체적으로 B.남성과 여성 그리고 B.Acc 사이의 시차연관구매가 가장 많이 발생하는 것으로 나타났다. 향후 시차연관구매가 많이 일어나는 브랜드에 대해 이후 구매할 가능성이 높은 브랜드에 대한 정보를 제공함으로써 구매를 유도하는 전략을 세울 수 있다

[표 3] 상세 시차 연관 Rule

Rule	고객수	Confidence	Support
BP ==> BE	53006	67.4	47.4
BE ==> BL	21355	22.4	19.1
BE ==> BP	15326	16.1	13.7
BL ==> BE	13249	29.6	11.8
BP ==> BL	9176	11.7	8.2
BL ==> BP	6531	14.6	5.6
BE ==> BJ	3024	3.2	2.7
BL ==> BJ	2664	6.0	2.4
BP ==> BJ	2561	3.3	2.3
BP ==>BE ==> BL	3653	6.9	3.3

4. 결론 및 향후 과제

고객과 시장은 항상 변화한다. 현재의 고객 분석 결과가 얼마나 유효한지는 모른다. 시장과 고객의 변화에 따라 고객 분석과 그에 따른 대응 전략은 항상 변화해야 한다. 데이터 마이닝 기법은 시장의 특성에 따라서 많은 영향을 받는다. 고객관리 방안 수립은 전사 기업전략 및 CRM 전략과의 연계 선상에서 이를 고려하여 수립되어야 한다.

본 논문에서는 단지 두개 브랜드간 연관 분석을 수행하였다. 여러 브랜드에 대한 연관성 분석과 상세하게 Segmentation 된 고객 등급별로 제품 구매 분석을 수행 해보다 정확한 차별화된 고객.판매 및 마케팅에 활용할 수 있도록 했으면 하는 아쉬움이 있다.

참고문헌

[1] (강현철 외)고객관계관리(CRM)를 위한 데이터마이닝 방법론[자유 아카데미]
 [2] C 패션 CRM 전략 산출물
 [3] (Bernd H. Schmitt) CRM 을 넘어 CEM 으로., 한언출판사
 [4] (홍태호, 전성용) 데이터마이닝을 이용한 고객이탈등급에 기반한 고객 세분화., 부산 대학교

[5] (송임영, 이태석, 신기정, 김경창) 고객관계관리를 위한 통합 데이터마이닝 모형 연구., 홍익대학교, 한국과학기술 정보원

[6] (정경용, 김중훈, 류중경, 임기욱, 이정현) 연관 마이닝을 이용한 고객 관계 관리 적용., 상지대학교, 대림대학, 선문대학교, 인하대학교

[7] (황현숙, 어운양) 연관 마이닝과 고객 선호도 기반의 인터넷 상품 검색 시스템 설계 및 구현., 부경대학교

[8] Srikant, R, Imielinski, T and Swami, A, "Mining Associations between sets of Items in Massive Databases," Proceedings ACM SIGMOD, 1993

[9] (조광현, 박희창) A Study of Association Rule Mining by Clustering through Data Fusion., 창원대학교

[10] (조장식, 이상식) Churn Predictive Model for University Students Using Data Mining., 경성대학교