

RFM 기법과 연관성 규칙을 이용한 개인화된 전자상거래 추천시스템

진병운*, 조영성**, 류근호***

Personalized e-Commerce Recommendation System using RFM method and Association Rules

Jin Byeong Woon*, Cho Young Sung**, Ryu Keun Ho***

요약

이 논문은 RFM 기법과 연관성 분석을 이용한 개인화된 전자상거래 추천 시스템을 제안한다. 제안된 전자상거래 추천시스템은 사용자의 평가 자료에 의존하지 않고 묵시적인(Implicity)방법을 이용하여 고객정보와 구매이력 정보를 기반으로 RFM(Recency, Frequency, Monetary) 기법을 이용한 고객 세분화와 교차판매(cross-sell)관계를 찾는 연관성 분석을 이용한 개선된 시스템이다. 또한 고객군별 구매특성 분석을 통하여 효율적인 마케팅 전략과 고객관계관리(CRM: Customer Relationship Management)방법을 제시한다. 현업에서 사용하는 데이터 셋을 구성하여 실험 및 평가를 통해서 효용성을 입증 및 평가하여 일대일 웹 마케팅을 실현하였다.

Abstract

This paper proposes the recommendation system which is advanced using RFM method and Association Rules in e-Commerce. Using a implicit method which is not used user's profile for rating, it is necessary for user to keep the RFM score and Association Rules about users and items based on the whole purchased data in order to recommend the items. This proposing system is possible to advance recommendation system using RFM method and Association Rules for cross-selling, and also this system can avoid the duplicated recommendation by the cross comparison with having recommended items before. And also, it's efficient for them to build the strategy for marketing and crm(customer relationship management). It can be improved and evaluated according to the criteria of logicity through the experiment with dataset collected in a cosmetic cyber shopping mall. Finally, it is able to realize the personalized recommendation system for one to one web marketing in e-Commerce.

• 제1저자 : 진병운 교신저자 : 류근호

• 투고일 : 2010. 02. 03, 심사일 : 2010. 03. 12, 게재확정일 : 2010. 09. 23.

*한국전자통신연구원 책임연구원 ** 동양공업전문대학 전산정보학부 산업체겸임 부교수

*** 충북대학교 전기전자컴퓨터공학부 교수(교신저자)

※ 이 논문은 2009년도 충북대학교 학술연구지원사업의 연구비지원에 의하여 연구되었음

▶ Keyword : RFM기법(RFM Method), 연관규칙(Association Rules), 교차판매(Cross-Selling), 추천시스템(recommend system), 전자상거래(e-Commerce)

연관규칙 및 순차패턴 탐사방법에 대해 알아본다.

I. 서론

초고속 인터넷의 보급과 인터넷의 폭발적인 대중화를 기반으로 다양한 형태의 전자상거래가 급성장하여 시장규모가 증대되고 있다. 또한 전자상거래에서 취급하는 아이템의 종류가 다양해져서 고객들이 원하는 아이템과 관련 정보를 효율적으로 추천하는 추천 시스템에 대한 연구가 활발히 진행되고 있다.[1,2] 고객들의 고객 성향을 분석하여 기업과 고객간의 상호 관계(customer relationship)를 높이고 고객에게 좀 더 많은 가치를 부여하는 개인화(personalization) 전략으로 옮겨 가고 있다.[3,4] 우수한 인터넷 상점들을 보면 대부분 고객에게 보다 개인화된 서비스를 제공하려는 노력을 전개하고 있다. 여기서 개인화란 고객 개개인을 대상으로 그에게 적합한 형태로 정보나 아이템을 제공해주는 서비스를 말한다. 이러한 시점에서 본 논문에서는 전자상거래(e-commerce)에서의 고객성향 분석이 가능한 고객 세분화의 RFM 기법과 교차판매(cross-selling)를 위한 연관규칙을 이용한 추천 시스템의 새로운 프레임워크를 제안한다. 또한 이를 통해서 기업 환경이 다변화하고 고객의 니즈가 다원화하는 시점에 기업의 비즈니스 전략 수립 및 고객중심 서비스 기반의 고객관계관리에 효과적인 방법을 제시하여 웹 원투원 마케팅의 실현을 가능하게 하였다. 본 논문의 구성은 다음과 같다. 제 II장은 관련 연구를 다루었으며 제 III 장에서는 제안한 제안 추천 시스템의 프레임워크를 제시 및 설명하며 제 IV장에서는 시스템 설계 및 구현하며 시스템의 실험적 분석 및 성능 평가를 실행한다. 마지막으로, 제 V장에서는 본 논문의 결론 및 향후 연구 과제에 대하여 기술한다.

II. 관련연구

기업들은 고객 데이터베이스를 이용하여 가치 있는 고객들을 발견해 내고 이들 고객에 대한 특성을 찾아냄으로써 보다 유용한 정보로 활용하여 수익을 증대시키는 마케팅 및 비즈니스 전략(business strategy)으로 고객 세분화 분석기법을 활용하고 있다.[5] 본 논문에서는 고객 세분화 RFM 분석 기법과 구매데이터를 기반으로 한 구매 패턴들 간의 연관성을 표현하는 여러 기법들 가운데 장비구니 분석과 관련이 깊은

2.1 고객 세분화(Customer Segmentation)

고객 세분화는 많은 고객들을 유사한 특성을 가지고 있는 고객들의 집단들로 분류하는 작업을 말한다. 고객세분화는 새로운 고객을 유치하고, 기존 고객을 유지하여 수익을 창출하는 것에 그 근간을 두고 있다 또한 고객 세분화는 STP(Segmentation Target Positioning)전략에서 선행되어야 할 중요한 과정이다. RFM 분석은 Recency, Frequency, Monetary의 약자를 따서 만든 명칭으로 여기서 R값은 최근 구매일자를 의미하며, F값은 일정기간 동안의 총 구매횟수를 그리고 M값은 일정기간 동안의 총 구매금액을 의미한다. RFM의 정의는 RFM의 세가지 요소를 가산 평균하여 단일 지표로 나타내고 이를 고객의 가치를 평가하는 지표로 삼는 방식이 바로 RFM에 의한 고객 점수 부여 방법이다.[6] RFM 분석은 고객 데이터를 세분화하기 위한 가장 간단하고 효과적인 방법의 하나이다.[7] RFM은 최근 구매월, 이용 빈도, 이용 금액 등 고객의 수익 기여도를 나타낼 수 있는 3가지 지표로 고객의 가치를 평가하는 방법이다.[8] 개개인의 라이프 사이클, 구매 수량, 구매 경향이 각각 다르기 때문에 이러한 고객의 경향을 분석하고 평가하기 위한 평가 방법이 필요하다. 이것이 바로 RFM 평가이다.[9] 우선 R값에 의해 5개의 세분화 세그먼트가 나누어지게 되고, 생성된 5개의 세그먼트는 F값에 의해 다시 5개로 나누어지며, 마지막으로 M값에 의해 각각의 세그먼트가 5등분되어진다. 따라서 전체 고객데이터베이스는 결국 $5 \times 5 \times 5 = 125$ 개의 세그먼트로 분할되어 진다. 고객의 가치를 평가하는데 가장 널리 사용되는 대표적인 RFM 점수 부여(Scoring) 시스템이다. 세 가지 요소를 기준으로 고객 각각에 대해 점수를 부여하고 세 가지 기준의 가중치를 주어 RFM 점수를 계산한다. 이 RFM 점수를 고객 가치를 평가하는 지표로 삼는 방식이 RFM에 의한 고객 점수 부여 방법이 라고 할 수 있다.[10] 고객 리스트를 점수순으로 정렬하여 고객을 세분화할 수 있다.

2.2 연관 규칙(Association Rule)

연관 규칙(Association Rules)은 사건들 사이의 상호 관련성 분석에 많이 이용된다. 연관성 측정방법은 어떤 특정문제에 대한 예측을 위해 상품 혹은 서비스의 거래에서 연관성이 많은 아이템 등을 분류하여 그것들과의 관계를 획득하는

것이다. 연관성 측정에서 얻어지는 연관성 규칙(Association Rule)은 다음 형식으로 나타낸다.

형식: If X Then Y (X → Y) (1)

이것은 데이터 셋에서 “X가 존재하게 되면, Y도 존재할 가능성이 높을 것이다”라는 연관성을 나타낸다. 그래서 특정 사건이 발생되면 연관 규칙에 의하여 다른 사건의 발생 가능성을 예측할 수 있다. 트랜잭션 R에 관련된 데이터 집합을 X, Y로 정의할 때 트랜잭션 R : X → Y 로 표현할 수 있다. 연관 규칙은 한 항목들의 그룹과 다른 항목들의 그룹 간에 강한 연관성을 나타내고 다음 식을 만족한다.

$X, Y \subseteq I, (X \cap Y) = \emptyset, Y \neq \emptyset$ (2)

이때 위의 식에서 X를 R의 조건부, Y를 R의 결과부라 한다. 만약 트랜잭션 R이 X를 지원한다면 Y도 특정한 확률에 의해 지원할 것이라고 예측할 수 있다. R에 의한 X와 Y의 관계성 예측을 연관 규칙이라 한다. 가장 널리 쓰이고 있는 지식발견의 형태 중 하나는 시장 장바구니 분석(MBA: Market Basket Analysis)이다. 이것은 사용자의 구매 경향을 파악하려는 곳에서 많이 쓰이는 방법이다. 장바구니 분석은 항목 집합(item set)들로 이루어진 데이터베이스에서 이들 항목들에 대한 관계성을 분석한다. 연관규칙을 찾기 위하여 일반적으로 지지도(support), 신뢰도(confidence), 향상도(lift)라는 척도를 사용한다. 고객의 구매패턴 파악 등 다양한 곳에 응용될 수 있으며, 온라인 전자상거래에서는 사용자로부터의 행동패턴을 분석하여 상품추천에 활용하고 있다. 연관규칙이 최근 추천시스템에 많이 적용될 뿐 아니라 데이터의 연관성을 찾아내는 연관규칙 알고리즘에 대한 연구도 진행되었다.[11]

- 지지도(support) : 생성된 연관규칙이 전체 항목에서 차지하는 비율을 뜻한다. 즉, 데이터베이스에 속한 전체 트랜잭션의 거래수 중 그 연관규칙을 지지하는 트랜잭션의 비율을 의미하며, 전체 거래 중 X와 Y를 포함하는 거래의 정도를 나타내는 식으로 표현된다.

$$P(X \& Y) = \frac{P(XUY)}{T} \quad (3) \text{sup}(XUY) = X \text{와 } Y \text{를 포함하는 거래수} / \text{전체 거래수}$$

- 신뢰도(confidence) : X를 포함하는 거래 중에서 Y가 포함된 거래의 정도를 의미하며 연관규칙의 강도를 의미한다. 전체부를 만족하는 트랜잭션이 결론부까지를 만족하는 비율을 말한다.

$$P(Y|X) = \frac{P(X \cap Y)}{P(X)} \quad \dots\dots\dots (4)$$

conf(R) = X와 Y를 동시에 포함하는 거래 수 / X를 포함하는 거래수

- 향상도(lift) : 규칙을 모를 때에 비하여 규칙을 알 때에 얼마나 향상되는가를 나타내고 있다.

$$\text{lift}(R) = \frac{P(Y|X)}{P(Y)} \quad \dots\dots\dots (5)$$

지지도는 연관규칙을 구성하는 항목집합을 포함하는 트랜잭션이 전체 트랜잭션의 몇 퍼센트나 되는지를 나타내는 것이고, 신뢰도는 규칙의 성립 정도를 나타낸다. 신뢰도가 규칙의 강도를 나타낸다면 지지도는 이 규칙이 전체 데이터베이스에서 가지는 통계적인 중요성을 나타낸다고 할 수 있다. 대표적으로 Apriori, FP-Tree, SETM, DIC, ARHP 알고리즘 등이 있다.

연관규칙을 찾아주는 알고리즘 중에서 가장 먼저 개발되었고 가장 많이 사용되고 있는 것은 Apriori 알고리즘이다. 이러한 동시 출현빈도를 찾는 과정은 데이터베이스에서 빈번항목을 추출하는 과정으로서 이 알고리즘은 두가지 단계로 구성된다. 첫 번째 단계 최소 지지도 설정 값에 따라 빈도수가 높은 항목의 집합들을 찾아내고, 이들 집합들 중에서 주어진 신뢰도를 만족하는 연관규칙을 찾아낸다. Apriori 알고리즘은 구매항목이 한 개인 빈번항목들을 먼저 구하고, 그 다음에 이것들을 이용해 구매항목이 두 개인 빈번항목들의 집합을 구하는 방식으로 수행한다. 다음은 <그림 1>은 Apriori 알고리즘을 이용한 연관 규칙 유추 과정 예를 나타낸 것이다.

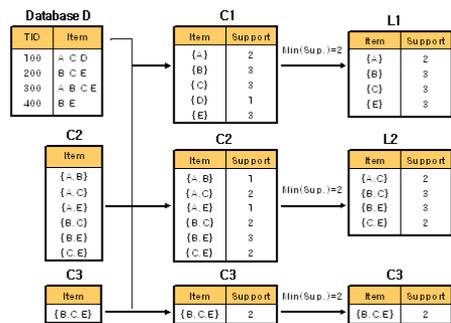


그림 1. Apriori 알고리즘을 이용한 연관 규칙 유추 과정 예
Fig. 1. example for routine of association rules using Apriori algorithm

2.3 순차 패턴(Sequential Patterns)

순차 패턴(Sequential Patterns)이란 연관 규칙에 시간 개념을 도입하여 트랜잭션에서 사용되는 데이터들로부터 특정한 시간대에 가장 많이 사용되는 데이터 시퀀스(sequence)의 패턴을 추출하는 데이터마이닝 기법이다. 연관 규칙과 유사하지만 트랜잭션을 통해 장바구니 데이터베이스로부터 고객이 선호하는 상품 항목(데이터 시퀀스)들의 순차 패턴을 추출하여 순차(sequence) 데이터베이스를 생성한다. 순차 패턴 기법은 1995년 Agrawal에 의해 처음 제안되었으며, 순차 패턴 기법은 트랜잭션의 발생 순서에 따라 각 트랜잭션에서 사용되는 상품 항목들의 연관성을 탐색하여 트랜잭션 간에 영향을 미치는 선행 항목을 검색한다. 즉 특정 항목 집합의 다음에 나타나는 항목 집합을 탐색하여 트랜잭션 간에 영향을 미치는 항목 시퀀스간의 패턴을 찾는다. 즉 고객의 상품 구매 패턴을 분석하여 향후 구매 가능한 상품 예측에서 사용된다. 다음 <표 1>은 단계별 순차 패턴 알고리즘의 기능을 나타낸 것이다.[12]

표 1. 단계별 순차 패턴 알고리즘의 기능

단계	세부 기능
1	정렬 단계 데이터 항목들의 정렬
2	항목 집합 탐색 단계 출현빈도가 높은 데이터 항목의 집합탐색
3	변환 단계 항목 집합들로 순차 DB의 생성
4	순차 패턴 탐색 단계 데이터 순차 패턴 탐색
5	최대빈도 탐색 단계 빈도수가 높은 순차 패턴 탐색

2.3 교차 판매(Cross-Selling)

교차판매란 아이템 판매시 상호 연관성을 고려하여 관리자가 미리 정한 지지도와 신뢰도를 바탕으로 아이템간에 도출해낸 연관 규칙을 이용하여, 구매를 유도하는 마케팅 전략이다. 이러한 교차판매는 마케팅 관점에서 웹 기반 시스템의 개인화를 지원하는 대표적인 데이터마이닝 기법이다.[10] 교차판매는 기존의 고객에게 새로운 아이템이나 서비스를 제안하는 프로세스이다. 기존 고객에 대한 구매 행태 데이터를 가지고 시작하게 되며, 대부분의 교차 판매는 단일 아이템 고객 획득을 위해 필요한 분석과 크게 다르지 않다. 중요한 것은 모든 고객에 대해 아이템 제안을 최적화하고 그렇게 함으로써 고객이 받은 제안이 구매자와 판매자 모두에게 최대의 이익을 줄 수 있다는 것이다. 데이터 마이닝 프로세스의 결과로 생성되는 모델은 새로운 제안을 수락할 확률을 예측하게 되며, 고객들의 예측된 확률을 통해 고객 등급을 결정함으로써 새로운 아이템에 대한 최적의 잠재고객을 확인할 수 있게 된다.

III. 제안시스템

3.1 제안시스템 구조

본 논문에서 제안하는 제안시스템의 구성도는 <그림 2>와 같으며 제안시스템의 절차 알고리즘은 <그림 3>과 같다. 먼저 로그인 사용자의 고객정보에서 Social data로 구성된 고객성향 분류코드와 고객 RFM점수를 인지한다. 그 후 로그인 사용자의 고객성향 분류코드를 이용하여 구매데이터에서 군집화(Clustering)된 집단(Cluster)을 구성한다. 고객 RFM 점수에 해당하는 구매데이터를 발췌하여 연관규칙을 적용하여 생성된 아이템의 선호도를 계산하여 선호도가 높은 아이템을 TOP-4로 추천한다. 다음 <그림 2>는 제안시스템의 구성도를 나타낸 것이다.



그림 2. 제안시스템의 구성도

Fig. 2. System configuration of the proposed system

전자상거래 환경에서 분석 에이전트는[13] 첫째 회원 가입을 통하여 고객집단을 분류 가능하도록 고객성향 분류코드로 부여하여 고객정보를 생성 및 관리한다. 둘째 고객정보에서 로그인 사용자의 고객성향 분류코드와 고객 RFM 점수를 인지한다. 또한 고객성향 분류코드를 이용하여 군집화(Clustering)된 집단(Cluster)을 구성하여 고객 RFM 점수에 해당하는 구매데이터를 발췌한다. 추천에이전트는 첫째 해당 군집에 속한 구매데이터를 기반으로 고객 RFM 점수를 적용하여 추천아이템을 생성한다. 둘째 장바구니 분석을 위한 연관규칙을 적용하여 교차판매를 위한 아이템을 구성한다. 셋째 아이템의 선호도를 계산하여 선호도가 높은 아이템을

TOP-4의 추천 아이템 목록을 생성한다. 이때 로그인 사용자의 구매 이력정보와 체크하여 중복 추천되지 않도록 한다. 학습에이전트는 Social data에 따른 인구통계학적 변수가 적용된 고객성향 분류코드를 갖는 고객정보 처리 및 관리하며 RFM 분석에 따른 학습 처리하여 고객 RFM점수 DB가 실시간으로 적용될 수 있도록 한다.

3.2 추천시스템 알고리즘

특정 고객 그룹별, RFM 분석에 따른 고객 RFM점수별로 구매촉진을 위해서 비즈니스 전략으로 구매 가능성이 높은 아이템을 추천하기 위해서 교차판매(cross-sell)를 위한 연관규칙을 이용한 연관성 분석 즉 장바구니 분석을 위해 고객 성향이 같은 고객 그룹 내의 구매 데이터를 기반으로 SQL을 통해 추천 아이템 생성이 가능하다. 본 논문에서는 묵시적인(Implicity)방법을 이용하여 우선 회원가입시 사용자의 social data를 통해 사용자 집단을 인구통계학적 변수 즉 나이, 성별, 직업, 사용자성향을 반영한 고객성향 분류코드 특성에 따라 세분화하는 과정을 통해 사용자정보를 고객성향 분류코드를 부여하여 생성 및 관리한다. 로그인 사용자는 사용자정보를 읽어 로그인 사용자와 같은 고객성향 분류코드를 갖는 사용자 정보내에서 구매이력 정보에서 자신의 속한 군집내의 고객성향 분류코드를 이용하여 군집화(Clustering)된 집단(Cluster)을 구성하고 해당 군집에 속한 구매데이터를 기반으로 고객 RFM 점수를 적용하여 추천아이템을 생성한다. 장바구니 분석을 위한 연관규칙을 적용하여 교차판매를 위한 아이템을 구성한다. 아이템의 선호도를 계산하여 선호도가 높은 아이템을 TOP-4의 추천 아이템 목록을 생성한다. 이때 아이템 추천시 로그인 사용자의 구매 이력정보와 체크하여 중복 추천되지 않도록 한다. <그림 3>은 RFM 기법과 연관규칙을 적용한 개인화된 추천 시스템 절차 알고리즘을 나타낸 것이다.

- Step 1* : 회원가입시 사용자의 social data를 통해 인구통계학적 변수가 적용된 즉 나이, 성별, 직업, 고객성향을 반영한 고객 성향 분류코드를 부여하여 고객정보를 생성 및 관리한다.
- Step 2* : 고객정보에서 로그인 사용자의 고객성향 분류코드와 고객 RFM 점수를 인지한다.
- Step 3* : 로그인 사용자의 고객성향 분류코드를 이용하여 구매데이터에서 군집화(Clustering)된 집단(Cluster)을 구성한다.
- Step 4* : 해당 군집에 속한 구매데이터를 기반으로 고객 RFM 점수를 적용하여 추천아이템을 생성한다.
- Step 5* : 장바구니 분석을 위한 연관규칙을 적용하여 교차판매를 위한 아이템을 구성한다.
- Step 6* : 아이템의 선호도를 계산하여 선호도가 높은 아이템을 TOP-4의 추천 아이템 목록을 생성한다.

Step 7 : 추천시 추천된 아이템을 로그인 사용자 구매 이력정보와 체크하여 중복 추천되지 않도록 한다.

그림 3. RFM 기법과 연관규칙을 적용한 개인화된 추천 시스템 절차 알고리즘

Fig. 3. Procedure algorithm for recommendation system using RFM method and Association Rules

IV. 시스템 구현 및 평가

4.1 실험 및 분석

본 시스템의 구현 및 실험 환경은 윈도우 운영체제 하에서 Apache/PHP/MySQL 웹서버 환경에서 사용하였고 버전은 다음과 같다.

- OS: Window XP
- Web Server : Apache HTTP Server Ver.1.3.34
- Database : MySQL Version 4.0.26
- Server-Side Application : PHP Ver.4.4.3

실험데이터의 목록으로 로그인 고객은 고객 회원정보를 읽어 자신이 속한 분류코드를 이용하여 구매정보에서 군집화(Clustering)된 집단(Cluster)인지하고 해당 군집화에 속한 고객 691명의 고객 정보와 화장품 아이템 200개를 대상으로 총 구매건수 3,444건의 구매데이터를 이용하여 로그인 고객의 군집성향에 맞게 구매데이터를 화장품 대분류 1개, 스킨케어 중분류 8개, 22개의 카테고리 소분류 200개를 아이템군별로 데이터를 취합하였다. 실험데이터의 구성은 현재화장품을 전문적으로 판매하는 인터넷 쇼핑몰 P사의 아이템 분류에서 대.중.소 분류로 실 쇼핑몰에서 사용되는 실 데이터를 중심으로 실험데이터 셋과 테스트 데이터 셋을 구성하였으며, 아래 <그림 4>는 모바일 환경에서 구현한 시스템 화면이다.



그림 4. 모바일을 이용한 시스템 구현 화면
Fig. 4. Capture the Screen of Proposed System using Mobile

4.2 실험적 데이터분석

본 논문에서는 실험 데이터를 통해서 추천을 위해서 고객 정보 및 구매정보를 이용하여 데이터 분석과 추천을 위한 RFM점수 분포를 분석 및 실험하였다. 다음은 구매이력데이터를 기반으로 한 RFM 점수 분포를 분석하여 현황을 나타낸 것이다. 본 제안 시스템에서의 고객점수는 고객 점수 테이블(point)에서 관리되며 점수 산출은 아이디어마다 부여된 필드의 종합 점수(score)로 계산된다. 최고 점수는 100점, 최하 점수는 0점이다. 다음은 전체 고객점수 691건을 고객별 RFM점수를 아래 산출식으로 나타내어 RFM점수별 고객 현황과 RFM 점수별 구매한 건수현황을 분석한 것을 나타낸 것이다.

① 점수 산출식은 아래와 같다.

$$RFM\ score = \{ (0.15 * R) + (0.35 * F) + (0.5 * M) * 20 \}$$

- R (최신성) : 제품의 구매일자를 기준으로 구입한 내역에 따라 5등급화 한다.

F (빈도성) : 고객이 구입한 전체적인 구매건수를 기준으로 5등급화 한다.

M (총구매액) : 고객이 구입한 제품의 총 구매액을 기준으로 5등급화 한다.

② RFM 점수별 고객 현황

다음은 총 691명의 RFM 점수별 고객 현황을 나타낸 것이다.

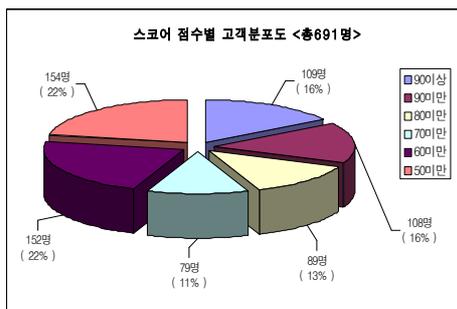


그림 5. 점수별 고객 현황
Fig. 5. The statistic of number of each customer score

위에서 <그림 5> 점수별 고객 현황에서 총 691명의 RFM 점수 중 60점 이하인 사람이 44%인 것을 알 수 있으며, 90점 이상은 16% 인 것을 알 수 있다.

③ RFM 점수별 구매현황

다음은 총 691명의 RFM 점수별로 회원이 구매한 건수를

나타낸 것이다.

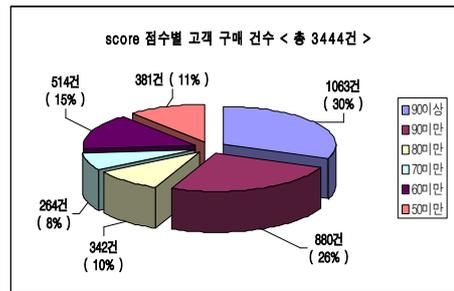


그림 6. 점수별 구매 현황
Fig. 6. The statistic of number of each purchase score

위에서 일반적으로 RFM 점수가 높은 고객이 많은 물건을 구입한 것을 알 수 있다. 구매이력데이터를 기반으로 각종 RFM점수별 분석이 가능하다. 고객별 RFM점수는 전체고객에 대한 RFM점수가 실시간으로 계산 및 관리되며 RFM별 구매현황은 RFM점수 가중치에 기반이 된다고 하겠다. 로그인 사용자와 같은 분류코드 내에서 로그인 사용자와 고객 RFM 점수가 유사한 고객이 구매한 아이템들을 Top-N으로 추천이 가능하며 RFM 점수별 현황을 참조하여 구매 촉진을 위해서 로그인 사용자와 유사한 고객이 구매한 아이템을 대상으로 같은 분류코드 내에서 RFM점수에 따른 고객이 구매한 아이템들을 연관규칙을 이용한 장비구니 분석을 통해서 구매 효율이 높은 아이템에 대한 교차판매를 위한 추천이 가능하다, 또한 RFM점수에 따른 아이템 추천의 경우에 연관규칙을 적용한 방법을 추가했을 때 추천 아이템의 정확률을 높일 수 있다. 다음은 SQL을 통해서 장비구니 분석을 실행한 결과를 정리한 예제이다.

```
mysql> select name, name2, count(*)
-> from two_tbl
-> group by name, name2
-> having count(*) > 1;
```

name	name2	count(*)
A.H.C 화이트닝 앵클멜라 피딘	로레알 센시 볼렌스 수딩 샴푸	3
A.H.C MeLa-fadin	찰론 링클리어 에멀전	4
A.H.C MeLa-fadin	청단홍단 컨디셔너	4
A.H.C MeLa-fadin	니백아 비디 아보카도오일	2
A.H.C MeLa-fadin	다빈치 헤어	2
A.H.C MeLa-fadin	비오메드 헤어테라피 샴푸	2
A.H.C MeLa-fadin	에워드하우스 수분크림	2
찰론 링클리어 에멀전	A.H.C MeLa-fadin	4
찰론 링클리어 에멀전	청단홍단 컨디셔너	4
찰론 링클리어 에멀전	니백아 비디 아보카도오일	2

그림 7. SQL을 이용한 장비구니 분석 결과
Fig. 7. The result of MBA using SQL

표 2 연관규칙을 이용한 교차판매를 위한 아이템목록

품목1	품목2	건수	지지도	신뢰도
~	~	~	~	~
A.H.C- Mela-fadin	참존 링클리어 에멀전	4	0.021	0.25
A.H.C- Mela-fadin	청단홍단 컨디셔너	4	0.021	0.25
~	~	~	~	~
A.H.C- Mela-fadin	다빈치 헤어	2	0.001	0.125
A.H.C- Mela-fadin	바이오메드 헤어테라피 CG샴푸	2	0.001	0.125
A.H.C- Mela-fadin	에뛰드하우스 수분크림	2	0.001	0.125
참존 링클리어 에멀전	A.H.C- Mela-fadin	4	0.021	0.25
참존 링클리어 에멀전	청단홍단 컨디셔너	4	0.021	0.25
참존 링클리어 에멀전	니베아 바디 아보카도오일	2	0.001	0.125
~	~	~	~	~
참존 링클리어 에멀전	바이오메드 헤어테라피 CG샴푸	2	0.001	0.125
참존 링클리어 에멀전	에뛰드하우스 수분크림	2	0.001	0.125
청단홍단 컨디셔너	A.H.C- Mela-fadin	4	0.021	0.25
청단홍단 컨디셔너	참존 링클리어 에멀전	4	0.021	0.25
청단홍단 컨디셔너	니베아 바디 아보카도오일	2	0.001	0.125
~	~	~	~	~
청단홍단 컨디셔너	에뛰드하우스 수분크림	2	0.001	0.125
~	~	~	~	~
~	~	~	~	~
~	~	187	~	~

4.3 제안 시스템 평가

고객 691명의 고객 정보와 화장품 아이템 200개를 대상으로 총 구매건수 3444건의 구매데이터의 제안 시스템에 대한 평가는 쇼핑몰을 이용해 본 경험이 있는 고객 691명의 고객 정보와 그리고 현장에서 사용하는 화장품 아이템을 대상으로 그들의 추천 3444건의 구매 데이터를 이용하여 로그인 고객의 고객성향에 맞게 구매데이터를 화장품 대분류 1개, 스킨케어 중분류 8개, 22개의 카테고리 소분류 200개를 아이템 군별로 데이터를 취합하였다. 실험은 특정 로그인 사용자와 고객성향이 같은 분류코드가 적용된 군집(Clustering)데이터 즉 Social data에 기반한 화장품 아이템 추천에서의 RFM 분석에 따른 RFM점수가 적용된 추천의 유효성과 추천된 아이템에 교차판매를 위한 연관규칙을 적용한 2가지 추천의 성능 평가를 포함하여 모두 세 방향으로 진행되었다. 우선 초기 화장품 아이템 추천의 유효성을 실험에 참가한 고객들의 구매데이터와 제시되는 화장품 아이템의 비교를 통해 이루어졌으며, 평가 방법으로는 정보검색 분야에서 보편적으로 사용되는 평가척도인 Precision을 응용하여 사용하였다.

Precision은 전체 추천된 상품 개수 중에서 사용자가 선호하는 아이템 개수의 비율로, 여기에서는 구매데이터에 기록된

아이템을 사용자가 선호하는 아이템으로 간주하였다.

$$precision =$$

$$\frac{\text{추천된 아이템과 일치하는 구매데이터 아이템 수}}{\text{Social data를 기반한 초기 추천 아이템 수}} \dots\dots\dots (6)$$

추천받는 대상이 되는 특정 로그인 사용자와 고객성향이 같은 social data를 통해 사용자 집단을 인구통계학적 변수를 갖는 분류코드(분류코드 : 3211)가 동일한 군집(clustering)내의 구매데이터를 이용한다. 로그인 사용자와 성향이 같은 분류코드의 Social data에 기반한 초기 추천 아이템을 분모로 하여 4가지 유형의 추천을 제시하여 평가 하였으며, 교차판매를 위한 연관규칙을 적용한 2가지 추천은 첫째 특정 아이템을 포함한 거래 중 두 아이템을 동시에 포함한 경우 비율에 따른 추천과 둘째로 전체 거래 중 두 아이템을 동시에 포함한 경우 비율에 따른 추천으로 구성하여, 실험유형별 추천조건에 따른 실험을 실시하였다.

조건 : 산업의 특성상 화장품 아이템을 고려한 실험을 대상으로 특정 로그인 사용자와 같은 분류코드 내에서의 실험을 대상으로 함

실험유형은 다음과 같다.

- A. 연관규칙이 적용되지 않은 추천
- B. 연관규칙이 적용된 추천1
- C. 연관규칙이 적용된 추천2

추천 조건은 다음과 같다.

- ①임의 추천
- ②최고 RFM점수를 가지는 고객들의 추천
- ③80점 이상의 RFM점수를 가지는 고객들의 추천
- ④50점미만 40점 이상의 RFM점수를 가지는 고객들의 추천

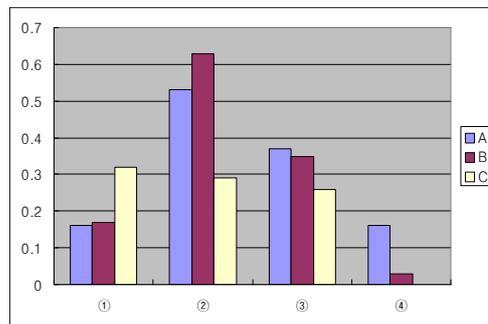


그림 8. 실험유형별 추천조건별 추천 평가결과 비교표
 Fig. 8. The comparison of result of recommending ratio for recommendation each experiment type and condition type

실험유형별 추천조건에 따른 실험결과는 실험유형 A의 연관규칙이 적용되지 않은 추천의 경우에 Precision 측정결과 추천 받는 대상에게 임의로 아이템을 제시하는 것보다 최고 RFM점수(②)를 갖는 고객들의 추천과 80점 이상의 RFM 점수(③)를 갖는 고객들의 추천에서 높은 추천결과가 나타났다. 고객 RFM 점수가 낮을수록 낮은 추천결과임을 알 수 있었다. 또한 연관규칙을 적용한 추천(1, 2)에서는 고객 RFM 점수가 최고 RFM점수(②)를 갖는 고객들의 추천과 80점 이상의 RFM점수(③)를 갖는 고객들의 추천에서 훨씬 높은 추천결과가 나타났다. 이는 고객RFM점수가 최고인 고객과 80점 이상인 고객에서의 추천에서 교차판매를 위한 연관규칙을 적용한 추천에서 더 좋은 추천결과를 얻을 수 있음을 입증함이라 하겠다. 특히 최고 RFM점수(②)를 갖는 고객들의 추천에서 구매촉진을 위한 교차판매를 시도하면 비즈니스 전략에서 특히 더욱 더 효과적일 수 있음을 입증하는 결과라 하겠다. 한편 50점미만 40점 이상의 RFM점수(④)를 가지는 고객들의 추천의 경우 연관규칙을 적용했어도 추천율이 낮은 결과로 나왔다. 이 경우 교차판매의 효과가 미미함을 알 수 있다고 하겠다. 본 제안시스템에서는 임의 추천보다 RFM분석을 통한 RFM점수에 따른 추천에서 좋은 결과를 얻었으며 특히 연관규칙을 적용했을 경우에 더 높은 추천결과를 얻을 수 있었다. 이는 특히 최고 RFM 점수의 추천의 경우에 교차 판매를 위한 비즈니스 전략 수립시 효과적인 결과를 얻을 수 있는 평가임이 입증한 결과라 하겠다.

V. 결론 및 향후 과제

극심한 경쟁 속에서 전자상거래 기업들이 살아남기 위해서 더욱 다양한 비즈니스 전략의 활용이 시급하다. 전자상거래에서의 비즈니스 전략은 고객에 대한 많은 가치를 부여하는 개인화(personalization) 전략으로 진이되고 있다. 본 제안 시스템은 다변화하는 시대에 구매 촉진 방법으로 고객 개인화를 바탕으로 구매 효율성과 추천의 성능을 향상시킬 수 있는 고객 세분화의 RFM기법과 연관규칙을 이용한 개인화된 추천 시스템을 제안한다. 다양한 고객성향에 대한 분석이 가능한 개인화 성향이 높은 social data의 인구통계학적 변수와 묵시적인(Implicit) 방법을 이용하여 고객정보와 구매데이터를 고객집단별로 군집화하고, 구매특성에 따른 RFM점수를 적용한 아이템 추천과 이들 아이템간의 교차판매 및 아이템 추천 효율을 위해 연관 규칙을 적용한 추천 알고리즘을 제안하였으며, 실험 및 평가결과를 통해 본 제안 시스템의 구매 효율성과 추천 효율성의 높은 결과로 추천시스템의 효용성을

얻을 수 있었다.

본 제안 시스템은 평가자료 기반의 사용자 프로파일 기반의 기존의 방법과 다르게 다양한 개인화 성향과 고객성향 분석이 가능한 RFM기법과 교차판매를 위한 연관규칙을 이용한 추천 시스템에 새로운 프레임워크를 제안함에 큰 의미가 있다. 또한 기업환경이 다변화하고 고객의 니즈가 다원화하는 시점에 기업의 비즈니스 전략 수립 및 고객중심 서비스 기반의 고객관계관리에 효과적인 방법을 제시하며 전자상거래를 위한 웹 원투원 마케팅의 실현을 가능하게 하였다. 향후과제로는 보다 다양한 산업분야의 실험 및 평가를 통하여 추천시스템의 유용성 테스트가 남아 있다.

참고문헌

- [1] 안광호, 김상용, 김주영. "인터넷마케팅원론", 법문사, 2001.
- [2] 장형욱, 한창엽 "CRM의최근 연구동향 및 향후과제", 경성대학교 경영연구논문지 제 21권 제1호, pp. 83-101, 2005
- [3] 김재경, 안도현, 조윤희, "고객선호도 변화를 반영한 상품 추천 방법론," 경영정보학연구 춘계논문집, pp.620-625, 2003.
- [4] 김종완, 오기욱, "e-CRM을 위한 개인화 마케팅 기법에 관한 연구", 한국컴퓨터정보학회 논문지 제 7권, 제 2호, 179-186쪽, 2002년. 1월.
- [5] 윤종욱, 윤종수, "균형적 고객세분화에 관한 사례 연구", 한국컴퓨터정보학회 논문지 제 11권, 제 2호, 303-317쪽, 2006년. 5월.
- [6] 박찬욱. "데이터베이스 마케팅-고객정보의 활용을 통한 기업경쟁력의강화", 연암사, 1996.
- [7] Weber, Alan., "A Simple way to use RFM, Target Marketing", Philadelphia, Mar, 1. 1997.
- [8] 김기서. "고객세분화 마케팅", 고원, 1999.
- [9] M Baier, "Elements of Direct Marketing", McGraw-Hill, 1983.
- [10] 김봉관. "데이터베이스 마케팅을 위한 고객관리", 남두도서, 2000.
- [11] 이기욱, 성창규, "데이터마이닝 기법을 이용한 추천시스템의 구현", 한국컴퓨터정보학회 논문지 제 11권, 제 1호, 293-300쪽, 2006년. 3월
- [12] 심장섭, "k-means 군집화와 순차 패턴 기법을 사용하는 VLDB 기반의 추천시스템 설계", 박사학위논문, 충북대학교 대학원 컴퓨터공학과, 2005

- [13] 최중민, “에이전트의 개요와 연구방향”, 정보과학회지, 7-16쪽, 제 15권, 제 3호, 1997년. 3월



류 근 호

1976년 : 숭실대학교 전산학과(이학사)
 1980년 : 연세대학교 공학대학원 전산전공(공학석사)
 1988년 : 연세대학교 대학원 전산전공(공학박사)
 1976년~1986년 : 육군군수 지원사 전산실(ROTC 장교), 한국전자통신연구원(연구원, 한국방송통신대 전산학과(조교수) 근무
 1989년~1991년 : Univ. of Arizona Research Staff (TempIS 연구원, Temporal DB)
 1986년~현재 : 충북대학교 전기전자 컴퓨터공학부 교수
 관심분야 : 시간 데이터베이스, 시공간 데이터베이스, Temporal GIS, 지식기반 정보검색 시스템, 유비쿼터스 컴퓨팅 및 스트림데이터처리, 데이터 마이닝, 데이터베이스, 보안, 바이오 인포메틱스

저 자 소 개



진 병 운

1983년 : 숭실대학교 전산학과(공학석사)
 2003년 : 충북대학교 전산학과 박사과정 수료
 1981년~현재 : 한국전자통신연구원 책임연구원
 2005년~현재 : 한국정보처리학회 지식및데이터공학연구회 위원장
 관심분야 : 시공간 데이터베이스, 데이터 마이닝, 기계학습, 프로젝트매니지먼트, 제약이론(Theory of Constraint), 품질경영



조 영 성

1989년 : 연세대학교 전산학과(공학석사)
 2008년 : 충북대학교 전산학과(공학박사)
 1982년~2001년 : 국세청 전산실, 미국 콘트롤데이터 한국지사(과장), 시나이미디어(부장), 유경제어기술구소(수석), 네오아이 엔씨(CEO)
 2002년~현재 : 가나소프트(대표), 동양공전 전산정보학부 산업체겸임/부교수
 관심분야 : 시공간 데이터베이스, 데이터 마이닝, 기계학습, 웹서비스, 웹 2.0, ebXML