

장바구니 분석을 이용한 인터넷 쇼핑몰에서 고객 행위 분석

Customer Behavior Analysis using Market Basket Analysis in Internet Shopping Mall

이병헌* · 김종우* · 이경미* · 한재룡*

* 충남대학교 자연과학대학 통계학과

Abstract

1. Introduction

인터넷을 통한 인터넷 쇼핑몰 시스템은 주로 고객을 위한 지능화된 서비스와 구매 의사결정지원에 중점을 두고 연구되어 왔으나, 인터넷 쇼핑몰 관리자 입장에서 마케팅 의사결정과 관련된 연구들은 아직도 미흡한 상태이다. 하지만 인터넷 쇼핑몰과 관련된 사용자 인증, 지불 및 결제, 보안등에 대한 표준화가 어느 정도 진행이 되면, 마케팅 의사결정 지원 기능이 인터넷 쇼핑몰 시스템의 핵심적인 성공 요인으로 자리잡을 것으로 보인다.

인터넷 쇼핑몰에서는 기존의 상점과는 달리 고객의 구매 이력 및 행동을 모니터링 할 수 있다. 따라서 고객의 다양한 취향을 파악하고 이를 마케팅에 활용함으로써, 기존 고객 및 잠재 고객을 대상으로 고객 유치 및 이탈 방지 효과를 얻을 수 있다. 이와 같이 고객의 행동을 분석함으로써 고객 만족도를 높이는 것은 물론, 기업은 마케팅에서 우위를 차지할 수 있는 효과를 얻을 수 있다. 인터넷 쇼핑몰에서 모니터링된 고객 행위 정보는 대용량 데이터의 분석을 필요로 하기 때문에, 전통적으로 고객의 특성을 파악하기 위해 널리 사용했던 RFM(Recency, Frequency, Monetary) 분석 방식이나 LTV(Life Time Value) 분석 방식보다는, 데이터웨어하우스를 기반한 데이터마이닝 기법의 활용이 필요하다.

2. Data mining

데이터마이닝 기법은 대용량의 데이터에 숨겨져 있는 데이터간의 관계, 패턴을 탐색하고 이를 모형화하여 업무에 적용할 수 있는 의미 있는 정보로 변환함으로써 기업의 의사결정에 적용하는 일련의

과정으로 정의된다. 대표적인 데이터마이닝 기법으로는 의사결정 나무(Decision Tree), 신경망 분석(Neural Network), 장바구니 분석(Market Basket Analysis), 링크 분석(Link Analysis) 등이 포함된다. 이 중 장바구니 분석은 연관성 규칙을 분석하기 위한 기법으로 교차판매(Cross Sales)에 활용되는 대표적인 데이터마이닝 기법이다.

장바구니 분석에서 연관성 규칙을 추출할 때, 수량화 된 기준이 필요하고, 그 기준은 *Support*, *Confidence*, *Lift* 세 가지가 있다. 만약 관심 있는 규칙이 “X 라는 상품을 구입한 사람은 Y라는 상품도 구입한다.”라고 한다면, 연관성 규칙의 세 가지 기준은 다음과 같이 설명되어진다.

(1) *Support*는 전체 거래 중 X와Y를 포함하는 거래량이 어느 정도인가를 파악하는 측정의 기준이다. (2) *Confidence*는 X라는 상품을 구매한 거래 중에서, Y라는 상품이 포함된 거래의 정도를 측정하는 기준이다. 이는 연관성 규칙의 강도를 나타낸다. (3) *Lift*는 임의로 Y를 구매하는 경우에 비해 X와의 관계가 고려되어 구매되는 경우의 비를 측정하는 기준이다.

의미 있는 연관성 규칙이 되기 위해서는 *Confidence*의 값과 *Support*의 값이 어느 정도 수준에 도달하여야 한다. 또한, *Confidence*와 *Support*가 자주 구매되는 항목에 대하여 높게 발생 할 수 있는데, 이런 경우 *Lift*를 고려하여야 한다.

3. Customer Behavior Monitoring

인터넷 쇼핑물의 특징 중의 하나는 전통적인 상거래에 비해서 고객에 대한 다양하고 방대한 양의 데이터를 추가적인 비용 없이 얻을 수 있다는 것이다. 즉, 인터넷 쇼핑물에서는 고객 행위들에 대한 데이터를 모니터링하고 축적할 수 있다. 실제로 웹에서의 사용자 모니터링에 대한 연구로, 카네기 멜론 대학의 Personal WebWatcher, IBM 도쿄 연구실의 Webby 등이 있다.

본 연구에서는 인터넷 쇼핑물 내에서 고객 행위 모니터링을 위한 기능을 서버릿과 자바 웹서버의 세션 트래커(Session Tracker)를 사용하여 구현하였다. 세션 개념을 통해 각 사용자에게 대한 연결의 지속성을 유지할 수 있고 다수의 페이지 요청들에 걸쳐 사용자의 신분과 상태를 유지할 수 있다. 구현된 모니터링 서버릿은 세션 트래커에 의해 제공되는 세션 객체를 최초 접근 요청 시에 생성하여 부여하게 되고, 세션 아이디를 응답 쿠키에 저장하여 웹 브라우저에게 전송한다. 쿠키에 저장된 세션 아이디는 사용자가 같은 웹사이트 상의 URL에 대한 서비스를 다시 요청할 때 웹 서버에 자동적으로 전달되며, 쿠키 속에 포함된 세션 아이디와 서버 안에 존재하는 세션 컨텍스트 안의 세션 아이디를 비교하게 된다.

또한 인터넷 쇼핑물 내에 모든 웹 다큐먼트에는 숨겨진 서식 항목(hidden form field)을 두어 지정된

행위에 대한 코드값과 부가 정보를 서버에 제공한다. 즉, 사용자가 어떤 작업이나 행위를 하였을 경우에 해당되는 숨겨진 항목이 모니터링 서버로부터 전달되어 로그 파일에 저장된다. 로그 파일에 저장된 고객 행위 데이터를 주기적으로 일괄 처리하여 로그 테이블에 저장한다. 이와 같이 모니터링 결과를 직접 로그 테이블에 입력하지 않고 로그 파일을 사용하는 것은 행위 모니터링으로 인한 인터넷 쇼핑몰의 성능 저하를 막기 위해서이다.

4. Algorithm과 Prototype

데이터마이닝 기법의 하나인 장바구니 분석을 활용하여 인터넷 쇼핑몰 내에서 상품의 구매, 웹 페이지 방문, 광고 클릭 등의 고객 행위를 분석하여 지식을 추출하기 위한 절차는 다음과 같다.

(1) 마케팅 의사결정자가 연관성 규칙을 정량화 한 기준 값(*Support*, *Confidence*, *Lift*)을 입력한다. (2) 분석에 포함시킬 고객의 행위의 범위를 선택한다. (3) 상점을 방문한 고객의 수(세션의 수)와 상점 내의 상품 페이지 개수를 계산한다. (4) 계산량과 추출된 규칙의 유용성을 위하여 주요 상품 페이지를 선정한다. (5) 상품 페이지 조합에 대하여 동일 세션에 포함된 개수를 계산한다. (6) 각 상품 페이지 조합에 대한 *Support*를 계산한다. (7) (1)에서 입력받은 *Support*값을 이용하여 상품 페이지 조합을 선택한다. (8) (7)에서 선택된 조합의 *Confidence*와 *Lift*를 계산한다. (9) (1)에서 입력받은 *Confidence*, *Lift* 값을 이용하여 연관성 규칙을 선택한다. (10) (9)의 결과 생성된 규칙에 대한 사후분석을 행한다.

프로토타이프 인터넷 쇼핑몰과 고객 행위 분석을 위한 장바구니 분석 알고리즘은 관계형 데이터베이스와 Java 기술을 활용하여 개발되었다. 즉, Microsoft Windows NT 4.0에서 Java Web Server, JDBC(Java Database Connectivity), 서버릿, Microsoft SQL Server 6.5 등을 사용하여 구현되었다. Java 환경과 JDBC를 활용한 것은 플랫폼 독립성과 데이터베이스 독립성을 갖추기 위해서이다.

5. Concluding Remarks

위에서 제시한 알고리즘을 활용하여 인터넷 쇼핑몰의 마케팅 관리자는 고객이 인터넷 쇼핑몰 방문시에 함께 방문하는 웹 페이지들에 대한 규칙들을 추출할 수 있다. 이러한 규칙들은 인터넷 쇼핑몰의 상품 전시, 추천 상품의 제시, 상품 패키징, 웹 페이지 간 Link 설계에 활용될 수 있다.

추후 연구 과제로는 추출된 마케팅 규칙들을 활용하여 고객에게 보다 지능적인 인터페이스를 제공하기 위해 방안에 대한 연구가 필요하다. 또한 고객 행위 분석을 위한 추가적인 데이터마이닝 기법에 대한 연구들이 계속적으로 필요하다.