

커뮤니티 점유시간을 이용한 상품추천 시스템의 설계 및 구현

천소영*, 윤성대**

*부경대학교 전산정보학과

**부경대학교 전자계산학과

*bada9245@empal.com

**sdyoun@pknu.ac.kr

A Design and Implementaion of Products Recommendation System Using the Residence time in Community

SoYoung Cheon*, SungDae Youn**

*Dept. of Computer and Information, Pukyung National University

**Dept. of Computer Science, Pukyung National University

요 약

전자상거래 시장에서 고객이 관심있는 카테고리가 포괄적이고, 우선순위를 알 수 없기 때문에 이 정보를 이용하여 상품을 추천한다는 것은 고객의 개인적인 취향을 충분히 반영하지 못한다고 할 수 있다. 이에 본 논문에서는 포털사이트에 가입된 기존의 고객이 활동하고 있는 커뮤니티들 중에서 빈번히 접속하고 장시간 머물러 있는 커뮤니티를 분석해 낸다. 또한 분석된 커뮤니티 카테고리와 연관된 상품을 고객에게 추천하도록 하여 구매율을 높일 수 있는 시스템을 설계하고자 한다.

1. 서론

인터넷 환경의 급속한 발달과 보급으로 인터넷 인구가 폭발적으로 증가함에 따라 전자상거래 시장이 급속하게 성장하고 있다. 또한 전자상거래 시장의 증가에 따른 사용자들의 다양한 요구에 대응하기 위해 개개인의 취향과 최근 관심사에 맞추기 위한 개인화는 전자상거래의 중요한 핵심기술로 떠올랐다. 현재 인터넷의 많은 전자상거래 업체들의 주된 목표는 기업이 가진 많은 상품들을 변화하는 사용자의 관심도에 맞추어 사용자에게 추천함으로써 추천 상품의 히트율을 증가시켜 고객의 만족도와 기업의 이익을 극대화하는 것이다.

상품추천을 위한 방법으로 내용기반 추천(Content Based)과 협업(Collaboration)을 통한 추천 등이 있다. 내용기반 추천은 정보검색 분야에서 그 기원을 찾을 수 있는 것으로 사용자가 과거에 구매를 하였거나 관심을 보인 상품의 프로필과 유사한 상품간의 비교를 통하여 추천이 이루어진다. 예를

들면 휴대전화를 많이 구입한 고객인 경우 새로운 휴대전화를 보여주고 책을 많이 구매한 고객에게는 책을 추천하는 방식이다. 이 방식은 과거에 고객이 구매한 내역을 바탕으로 추천이 이루어지므로 각 개인의 구매히스토리를 필요로 한다는 문제점이 있다. 협업을 통한 추천 방식은 고객이 구매한 상품들을 비교하여 다른 고객의 구매 패턴과의 유사도를 측정하여 유사한 구매를 보인 다른 고객이 많이 구매한 상품들을 추천하는 방식이다. 예를 들어 고객 A가 DVD와 책을 구매했고, 고객 B는 DVD와 CD를 구매했다고 할 경우 두 고객간의 유사도가 시스템에서 정한 임계치 이상이 된다면, 고객 A에게는 CD를 추천하고 고객 B에게는 책을 추천하는 방식이다. 이 방식은 시스템 구축 초기에 일정량 이상의 고객 데이터가 축적되어야 추천을 시작할 수 있다는 단점과 고객이나 상품이 추가될 경우 다른 모든 고객과의 유사도를 다시 계산해야 하기 때문에 고객이나 상품이 많아질 경우 성능에 문제가 생길 수 있다

는 단점이 있다[1][2].

본 논문에서 제안하는 방법은 포털사이트에 가입한 고객의 구매 내역이 없더라도, 커뮤니티 활동을 모니터링하고 분석한 뒤 분석된 내용에 대해 쇼핑몰 카테고리 및 연관성 규칙을 적용하여 클러스터링 하여 고객이 가장 활발히 활동하는 커뮤니티 카테고리 및 관련있는 상품을 고객에게 추천하는 방법이다. 이 방법은 변화하는 고객의 취향이 시간이 지남에 따라 바뀌더라도 그 변화에 맞춰 유동적인 추천이 가능하다는 장점과 고객자신의 커뮤니티 접근시간을 데이터마이닝하므로 다른 고객과의 유사도를 고려함으로써 생기는 성능저하 문제를 줄일 수 있는 장점이 있다.

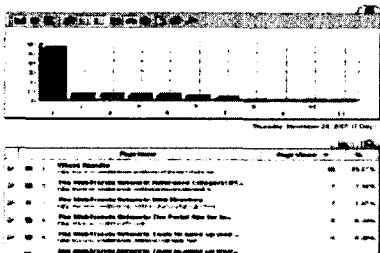
대부분의 고객이 가입초기에 채택하게 되어 있는 관심분야가 무의미하게 선택되거나 선택되지 않는다는 점과 선택을 한다 하더라도 관심카테고리의 우선 순위를 측정할 수 없는 부분에 대해 고객이 가입하여 접근하게 되는 커뮤니티페이지의 접근빈도와 시간을 측정하여 고객이 관심도를 반영하도록 하여 다음 로그인 시에 관련된 상품을 추천하도록 한다. 또한 각 커뮤니티와 전자상거래 상품이 카테고리별로 클러스터링 되는 특징을 이용하고, 가장 최근에 가입한 커뮤니티와 각 커뮤니티에 머무는 시간을 마이닝하여 사용자가 다음 로그인시 가장 관심있는 커뮤니티와 관련된 상품을 추천하는 방법이다.

본 논문의 구성은 2절에서 관련연구를 살펴보고, 3절에서는 제안하는 시스템의 각 에이전트에 대해서 설명한다. 이후 4절에서는 시스템 구현 부분에 대해 설명하고, 5절에서는 결론과 향후 연구 과제에 대해서 살펴본다.

2. 관련연구

2.1 통계적인 방법

웹 액세스 패턴을 추출하는 가장 단순한 형태는 통계적인 방법들을 이용하는 것이다. 예를 들어, 일



[그림 1] Web Trends를 사용한 예

정 기간 동안 가장 많이 액세스 된 페이지들의 순위를 알아내는 것과 같은 형태이다. 그림 1은 WebTrends라는 웹로그 분석 툴을 사용하여 고객이 액세스한 페이지를 시간대별, 페이지별로 분석한 통계정보를 그래프 형태로 본 경우이다.

2.2 연관규칙

웹 로그 데이터에 데이터마이닝의 기본적인 테크닉이라고 알려진 다양한 방법들을 적용하는 연관 규칙(association rule)은 웹 사이트의 페이지들 중에 자주 함께 참조되는 페이지들을 찾아내는 방법이다. 지지도와 같은 특정 값을 제시하고 이 값 이상의 확률을 가지고 함께 참조되는 페이지의 집합을 찾아낸다. 연관 규칙 중에 가장 기본적인 알고리즘은 Apriori 계열의 알고리즘을 발전시킨 많은 연구들이 제시되고 있다. 데이터베이스를 스캔하면서 빈발 항목 집합을 찾아내야 하는 연관 규칙의 경우 데이터베이스의 스캔 횟수를 줄이는 것이 중요한 관건이다. 이러한 연관 규칙은 주로 장바구니 분석(Market Basket Analysis)에 주로 사용되고, 또한 구매도 분석, 상품의 교차 매매(Cross-Marketing), 카탈로그 디자인, 열가 매출품(Loss Leader)분석, 상품진열, 구매 성향에 따른 고객 분류 등에 사용된다[4].

2.3 군집화(Clustering)

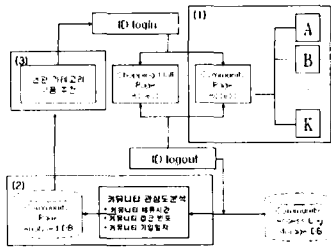
군집화란 유사한 성질을 갖는 항목들의 집합을 함께 그룹화하는 방법으로 이를 웹 액세스 패턴 마이닝에 적용하였을 경우, 비슷한 항목 패턴으로 웹 사이트를 그룹핑하여 이들에게 비슷한 추천 상품 사이트를 제공하는데 이용되고 있다. 일종의 문자열인 패턴 자체를 군집화한 뒤, 그룹별로 이를 가시화하는 연구들도 진행되고 있다. 이외에도 로그파일에 직접적인 프로세싱 작업을 통해서 순차적인 항목 패턴을 발견하는 방법들도 있다.

내용기반 필터링은 고객이 이전에 좋게 평가하거나 많이 구매한 아이TEM과 유사한 아이TEM 군을 추천 대상으로 정하기 때문에 전반적인 고객 군이 이질적인 평가를 보이는 상품 군의 추천에 높은 예측력을 보인다.

반면에, 협업필터링은 일반적으로 고객 군이 동질적인 평가를 보이는 상품 군에 대해 높은 예측력을 보인다. 그러나 이와 같은 방법들은 고객의 구매 이력과 아이TEM 선호도만 모델링 되므로, 고객의 개인적인 관심도를 반영하지 못한다는 단점이 있다[6].

3. 제안하는 추천시스템

본 논문에서 제안하는 시스템 에이전트는 커뮤니티 접속 모니터링 에이전트, 분석 에이전트, 상품추천 에이전트로 구성된다. 그림 2에서 (1)단계는 고객이 커뮤니티페이지를 액세스하는 것을 모니터링 하는 단계이고, (2)단계는 수집된 데이터를 분석하는 단계이며, (3)단계는 고객이 로그인 할 때 상품을 추천하는 단계이다.



[그림 2] 전체적인 시스템 구성도

3.1 모니터링 에이전트

그림 2의 (1)단계는 커뮤니티 단위의 점유시간을 측정하는 단계로서, 사용자가 로그인 한 후 접근한 커뮤니티페이지에서의 체류시간을 측정하여 데이터베이스에 저장한다.

커뮤니티코드와 연결한 시간정보를 세션에 저장했다가 커뮤니티페이지 이동시에 갱신정보를 저장한다. 예를 들어 고객이 A 커뮤니티에 접근하여 필요한 정보를 읽은 후 고객이 가입해있는 또 다른 B커뮤니티로 이동할 경우, 이전에 머물러 있었던 A 커뮤니티에서의 체류시간정보를 데이터베이스에 갱신한다. 이 과정을 커뮤니티를 벗어날 때 까지 반복한다. Community Access log를 저장하는 테이블의 형식은 그림 3과 같다.

회원번호	비밀번호	주요번호	이메일	전화번호	성	성명
회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인
회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인
회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인
회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인
회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인
회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인
회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인
회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인
회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인	회원(이)로그인

[그림 3] 커뮤니티 점유시간을 이용한상품추천시스템의 자료구조

3.2 분석에이전트

모니터링에이전트를 통하여 데이터 베이스에 저장된 커뮤니티 가입일자, 접근일자, 체류시간 정보를 사용하여 고객이 가장 활발히 활동하는 커뮤니티페이지를 분석한다. 식 (1)은 분석시에 사용할 패턴추출 알고리즘이다[7].

식 (1)에서 $C_{u,j}$ 는 사용자 u 의 Community에 대한 선호도를 나타내며, $F_{u,comm[j]}$ 는 사용자 u 가 Community j 에서 체류한 시간을 나타낸다. $F_{u,comm[j]}$ 는 사용자가 Community j 에 접속한 회수이다. $C_{u,d}$ 는 고객이 커뮤니티에 가입한 날짜를 나타내고 가중치를 적용하였다.

$$C_{u,j} = \frac{F_{u,comm[j]} * W_c + F_{u,comm[j]} * W_f}{\sum_{i=1}^n F_{u,comm[i]} * W_c + \sum_{i=1}^n F_{u,comm[i]} * W_f} \quad (1)$$

본 논문에서 제안하고 있는 방법은 고객이 가입하여 활동하는 커뮤니티 중에 가장 선호하는 커뮤니티를 분석해 내는 것이므로, 고객의 체류시간에 가장 높은 가중치를 주고, 다음으로 커뮤니티에 액세스한 빈도, 커뮤니티에 가입한 날짜순으로 가중치를 부여한다.

커뮤니티를 이동할 때마다 그 액세스한 시간과 체류시간을 기록하고, 로그아웃시에 이를 분석하여 고객의 커뮤니티 DB에 관심도를 반영한다.

3.3 상품추천에이전트

분석에이전트단계에서 분석된 가장 관심도가 높은 커뮤니티 카테고리를 바탕으로 고객이 다음 로그인 시에 상품 추천을 받을 수 있도록 한다.

커뮤니티에 가입하여 활동한 데이터 없는 고객의 경우에는 회원가입시에 선택한 관심분야를 "회원 관심 카테고리Table"에 등록을 하고 이를 토대로 관련된 상품을 추천하도록 한다. 그 외에 본 논문에서 목표로 하고 있는 고객인 커뮤니티에 가입하고 액세스한 로그 데이터가 있는 경우에는,커뮤니티 활동 DB 분석단계에서 계산된 카테고리를 기반으로 로그인시에 상품 추천을 받게 된다.

4. 시스템 구현

본 시스템에서는 고객이 회원가입시에 관심분야로 채택한 카테고리를 기반으로 상품을 추천하는 경우와 커뮤니티 관심도 계산에 의해 추천받은 상품을 비교해본다.

4.1. 구현 환경

본 시스템의 사용언어는 Java j2sdk1.4와 JSP이고, 웹서버는 resin-2.1.14를 사용하였으며, 데이터베이스는 MySQL을 사용하였다. 본 논문의 실험을 위한 데이터로는, 포털사이트의 카테고리 대분류코드와 중분류 코드를 사용하였다. 그리고, 고객의 커뮤니티 액세스 로그 기록은 10명의 고객이 30일동안 접속하여 활동한 커뮤니티 로그 기록을 임의로 입력하여 사용하였다.

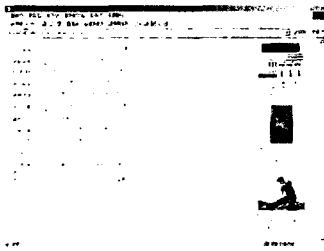
4.2 구현

표 1은 한 고객의 커뮤니티별 액세스 회수와 점유시간에 대해 식 (1)을 사용하여 선호도를 구한 예이다. 이렇게 구해진 값은 고객이 다음 로그인 시에 선호도가 가장 높은 카테고리나 관련있는 상품을 추천받게 된다.

활동여유치	카테고리코드	커뮤니티접속횟수	커뮤니티경유시간	선호도
dan9977	B1	27	12425	29.88
dan9977	A1	20	18912	32.09
dan9977	D1	8	5219	11.93
dan9977	C1	4	2884	6.39
dan9977	D2	4	4209	8.49
dan9977	A3	1	487	1.2
dan9977	B2	1	158	0.4
dan9977	F1	0	0	0
합 계		66	44295	

[표 1] 고객의 커뮤니티 액세스 로그 분석 예

[그림 4]는 고객이 로그인 한 후 선호도가 높은 카테고리나 관련이 있는 상품을 추천 받은 화면이다.



[그림 4] 로그인 후 상품추천받은 화면

5. 결론 및 향후 연구과제

기존의 추천 시스템에서는 협업을 통한 추천과 내용 기반의 추천방법으로 고객의 성향에 맞는 상품을 추천하도록 하였으나 이러한 방법을 적용하기 위해서는 고객 개인의 구매내역이 일정 수준이상 축적되어야 가능하다는 단점이 있다. 본 논문에서는 상품추천 성능의 향상을 위해 고객이 로그인 한 후, 각 커뮤니티페이지에 접근한 시간을 측정하여 가장 관심을 보이는 커뮤니티를 도출해내어 커뮤니티카테고리와 관련된 상품을 추천하는 방법을 제안하였다.

고객이 포털사이트에 가입할 당시의 채택한 관심분야를 바탕으로 상품 추천을 할 경우, 고객이 회원정보를 변경하지 않은 한 변화하는 고객의 관심분야를 반영할 수 없는 반면 본 논문에서 제안하는 알고리즘을 적용하면 고객이 최근에 가장 활발하게 관심을 보이는 커뮤니티 카테고리를 분석해내므로 그에 관련된 상품을 추천할 수 있다.

앞으로 선호도 계산 알고리즘을 더욱 보완하여 더욱 더 향상된 커뮤니티 선호 알고리즘을 적용하고, 관련 카테고리에서 최적의 상품을 선택해서 추천할 수 있도록 할 계획이다.

참고문헌

[1] Goldberg, D., nichols, D., Oki, B. M. and Terry, D., Using Collaborative Filtering to Weave an Information Tapestry, Communications of the ACM, 35(12): 61-70, 1992

[2] Good, N., Schafer, B., Konstan, J., Borchers, A.Sarwar, B., Herlocker, J.,Riedle J., "Combining Collaborative Filtering with Personal Agents for Better Recommendation", Proc. of the AAAI conference. p439~446,1999.

[3] Balabanovic, M., and shoham, Y., "Fab: Content-Based, collaborative recommendation", Communications of the ACM 40(30), pp66~72, 1997.

[4] Jiawei Han, Micheline Kamber "Data mining : Concepts and Techniques", Morgan Kaufmann, 2001.

[5] 박우창, 승현우, 용환승, 최기현(譯), "데이터 마이닝 개념 및 기법", 자유아카데미(出), 2003.9.

[6] 고병진, 유영훈, 조근식, "개인화 추천시스템을 위한 효율적 연관 규칙 방법" 2002년 한국정보처리학회 추계학술대회 2002.10.

[7] 장귀영, 조동섭, "사용자 프로파일을 이용한 웹페이지 추천" 2001년 한국정보과학회 춘계학술발표회 2001.4.

[8] 김춘호, 김준태, "카테고리 정보를 이용한 추천 성능의 향상" 2003년 한국정보과학회 춘계학술대회 2003.4.

[9] 이은영, 조동섭, "개인화 상품 추천을 위한 협력 필터링에이전트" 2001년 한국정보과학회 춘계학술대회 2001.4.

[10] 신헌섭, 이충세, "클러스터의 점유시간을 이용한 웹페이지 추천 기법" 2002년 한국정보과학회 추계학술대회 2002.10.