

구매자의 구매 패턴을 이용한 상품추천서비스에 대한 연구

신민수*, 황준원*, 김성학**, 이창훈*

*건국대학교 컴퓨터공학과

**유한대학 전자계산학과

e-mail : minshu@kkucc.konkuk.ac.kr

A Study on Product Recommendation Service using Purchasing Pattern of Buyer

Min-Su Shin*, Jun-Won Hwang*, Sung-Hak Kim**, Chang-Hoon Lee*

*Dept. of Computer Engineering, Kon-Kook University

** Dept. of Computer Science, Yu-Han College

요 약

대부분의 온라인 전자상거래에서 상품 추천 서비스는 사용자의 정보 또는 구매 이력을 가지고 카테고리 중심의 상품으로 상품을 추출하여 추천을 하는 구조이다. 또, 카테고리 중심의 추천을 하다 보니 단일한 구매 패턴에 의해서만 추천을 하게 되고, 상품에 각각에 대한 연관성을 찾아보기 힘들다. 또 단일 구매 패턴은 계산 비용이 작기는 하지만 사용자의 구매 패턴을 정확하게 반영하기 어렵다. 본 논문에서는 이러한 문제를 해결하기 위하여 카테고리 독립적이고, 다중 구매패턴을 고려한 상품 추천 서비스의 설계를 제안한다. 이를 위하여 단일 항목간의 구조화를 통하여 항목간의 연계성을 고려한 구조를 설계한다.

1. 서론

인터넷이 빠르게 보급되고, 사용자의 수가 많아지면서 전자상거래는 새로운 상거래의 형태로 빠르게 발전하고 있다. 빠른 변화에 맞추어 각 쇼핑 물들은 사용자에게 더 많은 정보를 제공하고, 편리한 사용자 인터페이스를 제공함으로써 보다 많은 수의 회원을 확보하려 노력하고 있다. 편리한 인터페이스 중의 하나는 상품을 추천해주는 추천 서비스이다. 이는 사용자가 쇼핑 물에서 구매한 정보, 행동 그리고 장바구니 등 사용자로부터 특정 행동 패턴을 잡아내어 분석하고 분석을 바탕으로 사용자에게 적합한 특정 상품을 구매하도록 유도하는 서비스이다.

현재의 인터넷 쇼핑 물에서 상품을 추천하는 방식은 각 카테고리별로 사용자의 관심도를 저장(Profile)하였다가 특정 카테고리에 접속하면 관심을 보인 카테고리의 상품을 사용자에게 보여주는 방식이다. 이는

사용자가 처음에 어떤 항목에 관심이 있는지 일일이 알려주어야 하며, 특정 카테고리의 상품에 대해서만 추천을 함으로써 다른 카테고리에 있는 상품에 관한 연관성을 찾아 볼 수 없다.

이에 본 연구에서는 상품들 간에 연계성 추출을 위하여 거래내역 데이터베이스의 마이닝(Mining)을 통하여 패턴을 추출해내는 기법을 사용하였다. 보통 마이닝을 통해서 추출해 낸 정보는 $A \rightarrow B$, 즉 “구매자가 A 라는 물건을 사면 B 라는 물건을 살 가능성이 높다”라는 패턴으로 추출을 해낸다. 이는 단일한 구매 패턴으로써 여러 품목에 관한 연관성을 찾기가 불가능하다. 따라서 본 연구에서는 다중 패턴 즉, $A \rightarrow B \rightarrow C$ 같은 패턴을 이용하여 이를 구조화하고 상품들 간에 연관성을 찾아내어 사용자에게 추천하는 시스템은 제안한다.

2 장에서는 상품 추천 시스템의 관련 연구와 데이

터베이스 마이닝 과정과 구조에 관해 논한다. 3 장에서는 추출된 패턴의 구조화와 상품들의 연계성에 관해 소개하고, 4 장에서는 실험, 5 장에서는 결론을 제시한다.

2. 관련 연구

이 장에서는 거래내역 데이터베이스로부터 상품 구매 패턴을 추출하기 위하여 필요한 데이터 마이닝의 기법 중에 본 논문과 관련이 깊은 연관규칙 (Association Rule)과 순차패턴(Sequence Pattern)에 대해 알아본다.

2.1 연관규칙(Association Rule)

데이터베이스에서 알려져 있지 않은 숨겨진 패턴을 탐사하는 연구 중에 연관 규칙에 대해 가장 많은 연구가 이루어 졌다. 연관규칙은 말 그대로 한 항목 그룹과 다른 항목 그룹 사이에 존재하는 강한 연관성을 찾아내어 그룹화 하는 클러스터링의 일종이다. 또한, 동시에 구매될 가능성이 큰 상품들을 찾아냄으로써 시장바구니 분석(Market Basket Analysis)에서 다루는 문제들에 적용할 수 있다. 연관규칙발견 알고리즘으로는 Apriori, OCD, SETM, DHP 알고리즘등 이 연구되었다. 연관규칙기법에 적용되는 데이터는 판매 시점에서 기록된 거래와 품목에 관한 정보를 담고 있다. 연관규칙 탐사과정은 크게 두 단계로 진행이 된다. 첫번째는 높은 지지도(Support)를 갖는 아이템의 집합을 식별하는 작업이고, 두 번째 단계는 높은 신뢰도(Confidence)를 갖는 연관규칙을 도출하는 작업이다. 여기서 지지도와 신뢰도의 개념은 아주 중요한 개념으로 빈발 항목 집합을 찾아내는데 있어 큰 역할을 한다.

지지도란 전체 트랜잭션에서 특정 패턴(A→B)이 차지하는 비율이고, 신뢰도란 A 를 구매하는 고객 중에 B 를 구매하는 고객이 차지하는 비율을 말한다 예를 들어 전체 거래건수 1000 건, A 는 500 건, B 는 300 건, A 와 B 는 250 건에 대해서 특정한 패턴(A→B)에 관해 신뢰도와 지지도를 구해보면,
 신뢰도(Support) : 250/1000 = 25%
 지지도(Confidence) : 250/500 = 50%

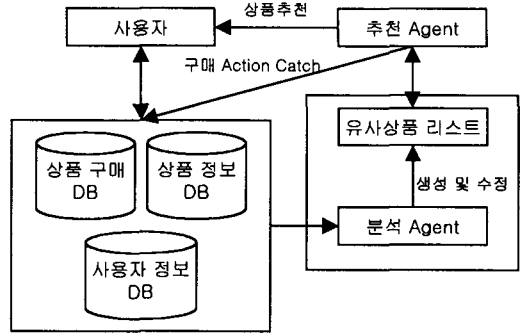
이를 해석하면 1000 건 중에 A 와 B 를 동시에 구매한 고객이 전체의 25%이고, A 를 구매했을 때 B 를 구매하는 고객은 25%중에 50%, 즉 전체의 12.5%를 차지한다는 것이다.

2.2 순차패턴(Sequence Pattern)

순차패턴은 동시에 구매될 가능성이 큰 상품 군을 찾아내는 연관규칙에 시간의 개념이 포함되어 순차적인 구매 가능성이 큰 상품 군을 찾아내는 방법이다. 순차패턴에서는 연관규칙 "A→B"는 상품 A 가 구매되면 일정 시간이 경과한 다음 상품 B 가 구매된다고 해석된다. 즉, 순차패턴은 구매 순서가 고려되어 상품 간의 연관성이 측정되고, 이에 따라 유용한 연관규칙

을 찾는 기법이다.

3. 시스템 구조



<그림 1> 상품추천 서비스의 구조도

사용자 쇼핑 물에서 구매를 하게 되면 구매 내역이 상품 구매 DB 에 남게 된다. 분석 Agent 가 이 정보를 이용하여 연관상품 리스트(Rule)을 추출하고, 구조화된 트리를 생성한다. 추천 Agent 는 생성된 연관상품 리스트를 이용하여 추천할 상품을 결정해서, 이를 사용자에게 추천하게 된다.

3.1 구매 패턴의 추출 및 구조화

추출된 패턴의 구조화는 다음의 4 단계를 통해 이루어진다.

- 1) 사용자 DB 에서 모든 사용자의 트랜잭션을 추출
- 2) 사용자의 트랜잭션을 모든 부분집합으로 분리
- 3) 생성된 단일 패턴의 분포도 계산
- 4) 분포도에 의거하여 구조화된 트리 생성

Rule 1: CPU→Cooler→Main Board→Memory Rule 2: Main Board →Memory→VGA Rule 3: CPU→Cooler→VGA→Memory Rule 4: CPU→Main Board→Memory Rule 5: Main Board→VGA
--

<표 1> 구매자의 원시 패턴

구매자의 원시 패턴을 연관규칙을 이용하여 단일 패턴으로 나눈다 단일 패턴으로 나누기 위하여 고려해야 할 조합은 예로 든 <표 1>의 Rule 1 같은 경우는 {CPU,Cooler}, {CPU,Main Board}, {CPU,Memory}, {Cooler, Main Board}, {Cooler,Memory}, {Main Board,Memory} 이렇게 6 개(4C₂) 이다. 이와 같은 방법으로 구매자의 원시 패턴의 모든 조합을 구한다. 이때, {CPU,Cooler}와 {Cooler,CPU}는 다르게 취급된다. 장바구니 분석에는 위와 같은 패턴이 같은 항목으로 취급 되지만, 제안된 패턴의 구조화에서는 패턴에 방향성이 고려되기 때문이다. 모든 패턴이 조합되었다면, 다음과 같은 표를 생성할 수 있다.

전체 거래건수 : 19 건에 대하여

CPU→ Cooler	CPU→ Main Board	CPU→ Memory	CPU→ VGA
2 (10.5%)	2 (10.5%)	3 (15.7%)	1 (5.2%)

Cooler→ CPU	Cooler→ Main Board	Cooler→ Memory	Cooler→ VGA
0 (0%)	1 (5.2%)	2 (10.5%)	1 (5.2%)

Main Board→ CPU	Main Board→ Cooler	Main Board→ Memory	Main Board→ VGA
0 (0%)	0 (0%)	2 (10.5%)	2 (10.5%)

Memory→ CPU	Memory→ Cooler	Memory→ Main Board	Memory→ VGA
0 (0%)	0 (0%)	0 (0%)	1 (5.2%)

VGA→ CPU	VGA→ Cooler	VGA→ Main Board	VGA→ Memory
0 (0%)	0 (0%)	0 (0%)	1 (5.2%)

<표 2> 각 항목간의 빈도수 테이블

이렇게 단일 항목으로 추출된 패턴들은 각 항목과 얼마만큼의 연관성을 가지고 있는지 나타내게 된다. 이 빈도수 테이블(support 5%이상)에 의거하여 트리를 구성한다. 트리를 구성함에 있어서 가장 빈도수가 높은 항목은 다른 상품과의 연관도가 가장 높다고 할 수 있다. 이 항목을 중심으로 트리를 구성해 나간다. 하위 노드의 확장은 빈도수가 가장 높은 항목을 첫번째 하위 노드 두 번째로 높은 항목을 그 다음 노드 이러한 방식으로 확장 노드를 생성해 간다.

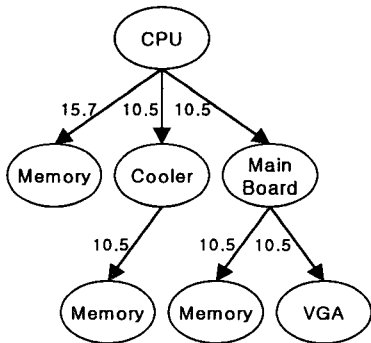


그림 2 구조화된 트리 (Support 5% 이상인 항목)

표 2에 따라 빈도수가 높은 항목을 우선으로 하여 하위 노드의 왼쪽에 두면서 점차 트리를 구성해간다. 이때, Support 5%미만인 연관 항목들은 빈도수의 저조에 의해 자동적으로 트리의 구조화에 반영되지 않는다.

다. 이와 같이 일정한 threshold를 두어서 많이 선택되지 않는 제품들을 걸러내는 작업이 필요하다. 이 값은 일정한 수치가 정해져 있는 것이 아니라, 통계치에 의해서 쇼핑물의 운영자가 선택해야 할 것이다.

3.2 추천 상품의 선택

상품 구매정보의 마이닝을 통해서 얻어진 연관 항목들에 대하여 빈도수 측정을 하여 특정 항목과 관련이 깊은 항목을 찾아 내었다. 이것을 빈도수 트리로 구조화 하여 연관 항목들에 대한 구조화를 수행 하였다.

추천 Agent는 고객의 구매 정보를 감시하다가 특정 항목을 구매하게 되면 그 즉시 구조화된 트리의 탐색에 의해 연관 항목을 추천하게 된다. 기본적으로 탐색은 pre-order 방식으로 Breath-First Search를 수행하게 되며, 탐색을 시작하게 될 Root Node는 고객이 처음 구매한 항목이 된다. 탐색된 하위 노드들 중에 Support 값이 가장 큰 노드를 추천하게 된다. 추천한 품목을 구매하게 되면 이는 다시 고객의 구매 정보 DB에 저장되며, 다시 트리의 탐색을 통해 다음 품목을 추천하게 된다. 만약 Root Node를 정할 때 트리에 여러 개의 동일 노드가 있다면 하위 노드의 Support 값이 가장 큰 노드를 Root 노드로 결정한다.

3.3 구조화된 트리의 수정

고객의 구매 데이터가 점점 쌓여 갈수록 사용자들의 구매 패턴을 트리에 반영해주어야 한다. 항상 최신 패턴을 반영해주기 위하여 고객의 구매 데이터가 특정한 Threshold를 넘을 때마다 트리를 수정한다.

예로 새로운 패턴에서 CPU→Cooler의 Support가 16.4%로 변경되었다면 CPU→Cooler에는 $|S_{old} - S_{new}|$ 를 추가하고 같은 레벨의 나머지 노드에는 $|S_{old} - S_{new}| / \text{동일 레벨 노드의 개수} - 1$ 만큼 각각 감소시켜 준다. 위와 같은 방식으로 트리를 점차 사용자의 구매 패턴에 맞추어 수정해 나간다면 사용자의 구매 패턴을 효과적이고 빠르게 추천 시스템에 반영할 수 있을 것이다.

3.4 도태된 항목의 삭제

시간이 지나면 지날수록 점점 많이 팔리는 항목이 있는가 하면 점점 팔리지 않아서 점차 구매 DB에서 찾아볼 수 없는 항목들이 생기게 된다. 이러한 항목들은 트리를 일정한 간격마다 한번씩 지지도가 특정한 Threshold를 넘지 않는 항목들을 삭제하면 점차 사라져가는 항목을 트리의 구조화에 반영할 수 있다.

4. 실험 및 고찰

가정: 구매자의 원시 패턴에 대해 단일 항목으로 분해하고 이를 일반적으로 사용하는 테이블구조와 구조화된 트리를 사용하여 추천했을 때, 테이블 구조보다 구조화된 트리가 항목간에 연관성을 더 반영해줄 수 있다.

- 1) 상품 구매 데이터 베이스에서 임의로 100건을 추출하여 각 항목간의 연관성을 테스트한다. 연관 테이블을 만들 항목은 지지도 6% 이상으로 작성하였다.

연관 항목	지지도
1A,2B	7.83%
1A,2D	6.79%
1D,2A	6.49%
2B,3A	6.57%
2D,3C	9.33%
3C,4A	6.57%
3C,4D	6.27%

<표 2> 단일 항목집합의 빈도수 테이블

- 2) 데이터 베이스에서 추출한 데이터에 근거하여 그림 3 과 같은 트리를 작성할 수 있다. 이 트리는 1A 항목과 나머지 2B, 3C, 4D 항목들 같이 연관도를 나타내고 있다.

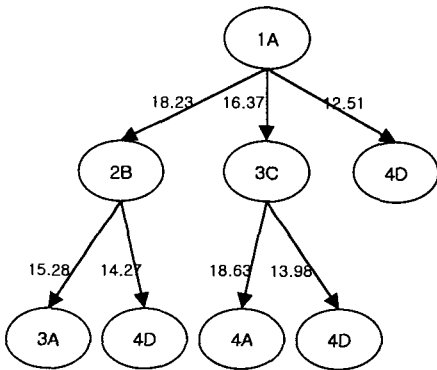


그림 3 구조화된 트리

단일 항목으로 연관 테이블을 작성하였을 경우, 1A 와 2B,2D 간의 항목은 연관 지을 수 있지만, 데이터에서 항목간에 연관도가 높은 3C 와 4D 에 관해서는 연관성을 추출해 낼 수 없다. 또, 마찬가지로. 2B 와 3A 는 연관 테이블로도 찾아낼 수가 있지만, 연관도를 무시할 수 없는 4D 는 2B 와 연관 지을 수 없다. 따라서 단일 항목간의 연관도 만으로는 다중 상품 간의 연관성을 결정지을 수 없다. 단일 항목 간이라도 구성 방법에 따라서 다량의 구매 데이터에서 숨겨진 연관성을 찾아 볼 수 있다.

5. 결론

본 논문에서는 인터넷 쇼핑 물에서 고객의 구매 패턴을 이용하여 상품을 추천하는 방법을 제안하였다.

이는 기존의 상품 추천에 비하여 다양한 상품군간의 간섭을 최대한으로 줄이면서 효과적으로 상품간의 연관도를 측정하여 고객에게 추천하여주는 시스템이

다. 기존의 쇼핑 물에서는 단일 항목간의 연관도만 반영하여 고객의 구매 데이터에 나타나있는 숨겨진 다중 항목간의 연관도를 밝혀낼 수 없었다. 그러나 제안된 구조에서는 각각 떨어져 있는 단일 항목 간에도 항목간의 구조화를 통하여 다중 항목간의 연관도를 밝혀냄으로써 조금 더 개선된 상품 추천 서비스의 구조를 제안하였다. 이를 위하여 고객 구매 데이터베이스의 구매 정보를 이용하여 일정량의 최근 구매 데이터가 생길 때마다 고객 선호도 트리를 수정한다. 고객 선호도 트리는 고객의 구매 데이터를 단일 항목으로 나누어 각 항목마다의 지지도를 구한다. 이 구해진 지지도를 기반으로 하여 구조화된 트리를 구성한다.

단일 항목간의 연계성을 가진 트리는 일반 단일 항목이 가지지 못하는 다중 상품들 간에 연계성을 표현할 수 있다.

향후 과제로는 본 연구의 확장으로서 각 개인의 구매 패턴을 추출하여 개인화 추천 서비스를 제공할 수 있는 구조를 연구하는 것과, 지금은 통상적인 방법으로 트리를 수정해주는 방법을 사용하였으나 개선된 트리의 수정을 위하여 인기상품 같은 단위 시간에 많이 팔리는 품목 같은 상품을 구조화 트리에 반영하기 위하여 가중치 필터링을 연구하는 일이다.

참고문헌

[1] J. Ben Schafer, Joseph Konstan, John Reidl "Recommender system in E-Commerce", GroupLens Research Project.

[2] Nathaniel Goods, J.Ben Schafer, Joseph A.Konstan, Al Borchers, Badrul Sarwar, Jon Herlocker, and John Riedl "Combining Collaborative Filtering with Personal Agents for Better Recommendations", GroupLens Research Project

[3] Emmanue Nauer, Jacques Ducloy, Jean-Charles Lamirel, "Using of multiple data source for information filtering: fist approaches in the MedExplore project."

[4] 김종우, 김영국, 유관중, 강태근, 한상혁, 전혜성, "비감독하의 학습을 이용한 전자상거래 시스템에서의 개인화된 광고 제공"

[5] 이광형, 김정재, 오해석 " 인터넷 쇼핑물을 위한 지능형 대화기반 판매 에이전트 시스템"

[6] 이치훈, 정준, 정대진, 김용환, 이필규, "협력 및 내용기반 정보 여과 방식을 이용한 추천 시스템 설계", 제 14 회 산.학.연 멀티미디어 산업기술 학술대회 (1999.11)