

사용자의 묵시적인 정보를 이용한 추천 시스템

정준⁰ 김용환 이필규
인하대학교 전산학과
{g1991263, g1991263@inhavision.inha.ac.kr}

A Recommender System using Implicit Information of Users

Jun Jeong⁰ Yong-Han Kim Phill-Kyu Rhee
Dept. of Computer Science and Engineering, Inha University

요약

인터넷의 발달로 인해 사용자들은 이제 인터넷에서 필요한 정보를 습득할 수 있을 뿐만 아니라, 생활에 필요한 여러 가지 활동들을 할 수 있게 되었다. 그 중에서 많은 관심을 갖는 것은 구매 활동이다. 따라서 수많은 기업들이 사람들의 구매 활동에 대응하여 전자상거래에 투자를 하고 있고, 현재는 Amazon.com과 같은 세계적인 사이트도 나타나기 시작하고 있다. 또한, 전자상거래 사이트들은 사용자들의 구매 활동을 도와주기 위해 추천 시스템의 도입을 추진하고 있다. 추천 시스템은 사용자들로부터 얻어진 정보를 학습하여 이용 가능한 상품 중에서 고객이 좋아할 만한 것은 추천해주는 시스템이다. 본 논문에서는 명백하게 사용자에게 정보를 요구하는 방법 대신에 묵시적인 정보 즉, 구매 활동에서 발생하는 정보를 이용한 음악 추천 시스템을 제안하고자 한다.

1. 서론

인터넷의 발달은 사람들의 생활 방식을 바꾸어 놓을 만큼 영향력을 행사하고 있다. 전화번호 검색과 같은 단순한 정보를 획득뿐만 아니라 오락 산업, 전자상거래 등 기존의 생활 방식을 인터넷으로 전환이 가속화되고 있다.

그 중에서 전자상거래는 많은 관심을 받고 있으면 막대한 투자도 이루어지고 있다. 따라서 전자상거래를 목표로 하는 사이트간의 경쟁도 치열해지고 있으며 사용자들의 관심을 형성하기 위해 많은 노력을 하고 있다.

이런 상황에서 구매 촉진을 위한 추천 시스템이 도래되기 시작했으며 Amazon.com과 같은 일부 선진 사이트들은 기본적인 서비스를 시작하고 있다. 여기서 말하는 추천시스템이란 사용자로부터 명백하게 혹은 묵시적으로 정보를 획득하여 학습을 통해서 사용자가 선호할 만한 상품을 추천해 주는 시스템이다.[9] 이것은 구전이라는 입에서 입으로 전해지는 사회현상을 모델링한 시스템이다. [6]

본 논문에서 구매 활동에서 발생하는 정보와 사용자 디렉토리 정보와 같은 묵시적인 사용자 정보를 통해서 음반 판매 사이트에서 활용될 수 있는 음악 추천 시스템을 제안하고자 한다.

2. 관련 연구

추천 시스템에 관련된 연구는 세계적으로 많이 이루어지고 있다. [5][6][9] 주요한 연구분야는 사용자 정보 습득 방법, Rating 방법[8], 협력적 여과 알고리즘[3], 내용 기반 여과 알고리즘 등이 있다.

2.1. 사용자 정보의 습득

사용자 정보의 습득은 추천시스템에서 중요한 문제이다.[2] 일반적인 정보의 습득과 마찬가지로 정보의 종류와 크기, 습득 시기와 방법 등을 고려해야 한다. 추천시스템에서는 대규모의 사용자로부터 선호도 정보를 수집한다. 초기의 추천시스템에서 주로 Rating 정보를 이용했다 [4][6]. 즉, 사용자가 좋아할만한 아이템의 리스트를 제공해서 각 아이템을 선호도에 따라서, 5단계 혹은 7단계로 구분하여 평가를 하도록 했다.[4][6] 그러나 사용자가 직접 선호도를 평가하는 방법은 평가의 정확성은 확보할 수 있으나 사용자에게 부담을 주는 결과가 되었다. 그래서 묵시적인(implicit) 평가 방법에 대한 연구가 진행되었다.[8] 묵시적인 평가 방법에 주된 동기는 아이템을 평가하는 사용자에 대한 비용을 제거하는 것이다. 비록 묵시적인 평가 정보를 처리하고 저장하는데 필요한 계산적인 비용이 발생하지만 이것은 사용자에게 가려져 있다. 묵시적인 평가 정보로 사용할 수 있는 자원들은 [8]에서 제안하고 있다. 예를 들면, 구매, 선택, 저장, 삭제 등과 같은

사용자 행위 정보를 기반으로 한다.

2.2. 협력적 여과 알고리즘

추천시스템은 추천할 아이템을 선택하기 위하여 협력적 여과(Collaborative Filtering) 알고리즘을 사용한다. 협력적 여과란 사회적 여과(Social Filtering)라고도 하며 유사한 기호를 가지는 다른 사람들의 선호도에 기반해서 아이템을 여과한다.

협력적 여과는 [7]에서 처음으로 제안했으며 기존의 내용 기반 여과의 문제점을 해결하고자 제시되었다. 즉, [6]에서 언급하듯이 내용 기반 여과에서 아이템은 기계가 파악할 수 있는 형태이어야 하며, 속성들은 손수 아이템에 할당되어야 하고, 우연한 발견을 형성하기 위한 어떠한 방법도 가지고 있지 않다.

협력적 여과에서 사용할 수 있는 알고리즘들은 기계학습, 통계적 방법, AI 등의 분야에서 파생되었으며 알고리즘들의 체계적인 분류 및 특성들은 [3]에서 잘 언급되고 있다. [표 1]은 간략하게 알고리즘의 분류를 나타내고 있다.

메모리 기반 알고리즘	Correlation
	Vector Similarity
메모리 기반 알고리즘 확장	Default Voting
	Inverse User Frequency
	Case Amplification
모델 기반 방법	Cluster Models
	Bayesian Network Model
	Case Amplification

[표 1] 협력적 여과 알고리즘 분류

3. 음악 추천 시스템

3.1. 목시적인 평가 정보

본 논문에서 제안하고자 하는 추천시스템은 음반 판매 사이트에서 사용자의 구매 활동에서 발생하는 정보와 사용자의 디렉토리 정보와 같은 목시적인 정보를 이용한 음악 추천 시스템이다. 즉, 구입한 음반 정보(구매 정보), 구입하지는 않았지만 구입하려고 고려되었던 정보(장바구니 정보)와 사용자가 직접 입력한 디렉토리 정보를 선호도 정보로서 이용한다. 모아진 목시적인 선호도 정보는 명시적인 선호도 정보로 변환된다. 명시적인 선호도 정보는 5단계로 나타내고 0에서 1사이의 값으로 정규화가 된다. 정규화가 된 선호도 정보는 협력적 여과 알고리즘의 학습 자료가 된다.

3.2. 추천 알고리즘

본 시스템에서 사용한 협력적 여과 알고리즘은 [6]에서 제안한 Constraint Pearson's r 방법을 사용한다.

상관계수인 r은 가중치를 구하기 위한 기초가 된다. 사용자 x와 아이템 y 사이의 상관계수는 다음과 같이 정의된다.

$$B_{xy} = \frac{\sum(U_x - M)(U_y - M)}{\sqrt{\sum(U_x - M)^2 \sum(U_y - M)^2}} \quad (1)$$

여기서, B_{xy} 는 U_x 와 U_y 사이의 "correlation"의 새로운 측정이고 M은 평가 범위의 중간값이 된다. B의 크기는 1보다 항상 작다. 이것은 Schwarz Inequality를 사용해서 증명된다.

식 (1)을 이용해서 모든 사용자간의 B값을 구한 후 어떤 한계값 L이상의 값을 갖는 사용자들은 유사한 사용자라고 인정되어지면 유사한 사용자들간의 가중치는 다음과 같이 정의된다.

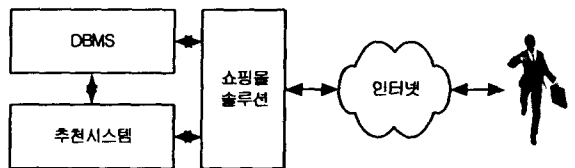
$$W_{xy} = \left(\frac{B_{xy} - L}{1 - |L|} \right) \quad (2)$$

식(2)에서 구해진 가중치를 이용해서 사용자 X의 아이템 Y에 대한 예상 선호도는 다음과 같이 정의된다.

$$P_{xy} = \bar{U}_x + \frac{\sum_k^N W_{ky} \times S_{ky}}{\sum_k^N W_{ky}} \quad (3)$$

S_{ky} 는 사용자 k가 아이템 y에 대해서 평가한 값이고 \bar{U}_x 는 사용자 x가 평가한 값들의 평균이다.

3.1. 전체적인 시스템 운영 환경



[그림 1] 전체적인 시스템 운영 환경

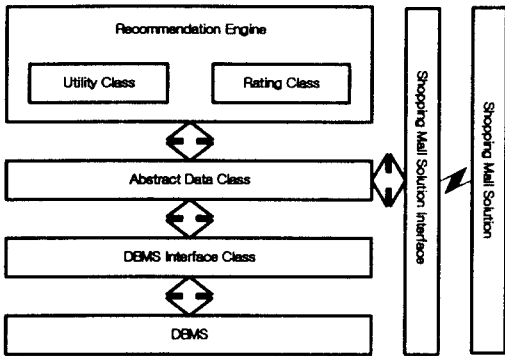
본 시스템의 전체적인 운영환경은 [그림 1]에서 보여진다. 대부분 전자상거래 사이트들은 쇼핑몰 솔루션을 사용하고 있다. DBMS는 쇼핑몰 솔루션이 사용하는 DBMS와 별도로 추천시스템을 위한 DBMS로서 쇼핑몰 솔루션이 제공하는 정보 및 추천시스템에 관련된 정보를 관리한다. 따라서, 추천시스템은 쇼핑몰 솔루션이 제공하는 정보의 복사와 추천리스트의 생성과 쇼핑몰 솔루션이 사용할 수 있도록 추천리스트를 가공한다. 가공된 추천리스트는 쇼핑몰 솔루션의 DBMS에 유지된다.

3.2. 세부 시스템 구조

[그림 2]는 세부 시스템의 구성을 보여준다. 본 추천 시스템은 15개의 클래스와 9개의 DBMS 테이블로 구성된다. 클래스들은 테이블에 대한 추상 클래스와 추천엔진에 관련된 클래스들로 분류될 수 있다. [그림 2]는 세부 시스템의 개념적인 다이어그램이다.

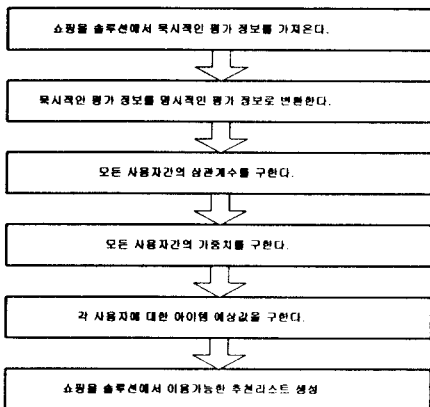
DBMS Interface Class 는 추천엔진이 사용되어지는 DBMS에 대한 독립성을 제공하여 준다. 예를 들면 JDBC나 ODBC 드라이버 등이 있다. Abstract Data Class 는 DBMS 의 테이블에 대한 추상 클래스로 추천엔진에게 테이블 연산에 대한 자세함을 감추고, DBMS Interface Class 를 통해서 실제적인 연산을 수행한다.

Utility Class 는 식 (3)에서 사용자의 평가값의 평균을 구하는 것과 같은 추천엔진을 지원하기 위한 부수적인 작업을 전담한다. Rating Class 는 구매자료와 같은 쇼핑물 솔루션이 제공하는 정보를 평가하여 추천엔진에게 제공하고 추천엔진이 생성한 리스트를 쇼핑물 솔루션이 사용 가능한 형태로 변환하는 역할을 한다.



[그림 2] 세부 시스템 구성도

Recommendation Class 는 추천리스트를 생성하기 위한 핵심적인 역할을 한다. 3.2절에서 언급하고 있는 알고리즘에 대한 구현이다. 즉, 모든 사용자간의 상관계수를 구하고, 그 수치를 이용해서 가중치를 구한다. 마지막으로 각각의 사용자에 대한 추천리스트를 생성한다.



[그림 3] 추천리스트를 생성하기 위한 전체적인 과정

[그림 3]은 추천리스트를 생성하기 위한 전체적인 과정을 설명하고 있다.

4. 향후 연구 방향 및 결론

추천시스템이 사용하는 협력적 여과는 a) 새로운 아이템이 평가되지 않는다면 추천 불가, b) missing data : 평가하지 않은 아이템, c) 추천 대상의 내용을 참고하지 않음 d) 평가값의 회소성 등과 같은 문제점을 나타낸다. 따라서 상기 문제점을 해결하기 위해서 다각적인 연구가 필요로 한다.

추천시스템은 전자상거래의 활성화로 구매 촉진을 위한 중요한 요소를 작용할 것이 분명하다. 그러므로 추천시스템에 필요한 기술의 확보 및 새로운 알고리즘 개발이 요구되어지고 있다.

5. 참고문헌

- [1] Pattie Maes, "Agents that reduce work and information overload", Communication of the ACM, Vol. 37, No 7, pp. 31-40, 1994
- [2] Nathaniel Good et al., "Collecting user access patterns for building user profiles and collaborative filtering", In AAAI'99, pp. 439-446, 1999.
- [3] John s. Breese, David Hecherman, and Carl Kadie., "Empirical Analysis of Predictive Algorithms for Collaborative Filtering", In Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence(UAI-98), pages 43-52, San Francisco, July 24-26 1998.
- [4] Paul, R, Neophytos, I, Mitesh, S. Peter, B, John, R(1994)., "GroupLens : an open architecture for collaborative filtering of netnews", In Proceedings of ACM CSCW'94 Conference on Computer Supported Cooperative Work, pages 175-186, 1994
- [5] David Maltz and Kate Ehrlich., "Pointing the way: active collaborative filtering", In Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence(UAI-98), pages 43-52, San Francisco, July 24-26 1998.
- [6] Upendra Shardanand and Pattie Maes, "Social Information Filtering Algorithms for Automating Word of Mouth", In Proceedings of the CHI'95. Denver, CO. May 1995
- [7] David Goldberg, David Nichols, Brian M. Oki, and Douglas Terry, "Using collaborative filtering to weave an information tapestry", Communication of the ACM, 35(12):61-70, December 1992.
- [8] Nichols, D. M. 1997., "Implicit Ratings and Filtering", Proceedings of the 5's DELOS Workshop on Filtering and Collaborative Filtering, 10-12. Budapest, Hungary, ERCIM.
- [9] Resnick, P. and Varian, H. R., "Recommender systems", CACM. 40(3), pp. 56-58, March 1997