

연관규칙 마이닝을 이용한 개인화된 추천시스템

성창규⁺·류길수⁺⁺·김태진⁺⁺⁺

Personalized Recommend System Using Mining for the Association Rule

Chang-Gyu Sung⁺, Keel-Soo Rhyu⁺⁺ and Tae-Jin Kim⁺⁺⁺

Abstract : Recommend Systems are being used by an ever-increasing number of E-Commerce to help customers find products to purchase. Recommend Systems offer a technology that allows personalized recommendations of items of potential interest to users based on information about similarities and dissimilarities among different customers tastes. In this paper, we design and build a Recommend System using the historical customer movie purchase transactions and extracts the knowledge needed to make association recommendations to new customers.

Key words : Personalization, Recommend System, Data Mining, Association Rule, KDD

1. 서론

현재 여러 분야에서 데이터의 엄청난 양의 증가율로 “정보의 홍수” 라고 할 정도로 대규모의 데이터를 이루고 있다. 하지만 데이터를 수집하는 능력에 비해 대규모의 데이터를 분석하고 이해하는 능력은 훨씬 떨어진다. 데이터의 가치는 “얼마나 많이 보유하느냐” 하는 것이 아니라 얼마나 빠르게 효과적으로 데이터를 축소하고 분석하여 처리하는 것이 중요하다. 마케팅 분야에서 최근 가장 인기 있는 주제는 소비자의 취향과 감각을 지능적으로 분석, 다음번에 구매할만한 상품을 제때 추천하는 것이다.[3] 인공 지능의 기계 학습 분야에서 출발한 개인화 추천 시스템은 웹의 인기를 타고 널리 사용되고 있다.[6] 예를 들어 인터넷 서점인 아마존(Amazon.com)은 비슷한 성향의 사용자들이 이미 구매한 상품을 자동적으로 추천하는 시스템을 갖추고 있다.[3][4]

실제적으로 개인화가 효과를 거두려면 실시간 추천

시스템을 갖추어야 한다. 추천 시스템의 구성 요소 중에서 중요한 추천 엔진에서의 고객의 성향 분석을 하는 방법은 아직까지 그 분류 기준 조차도 사람에 따라 다르지만 일반적으로 많이 사용하는 방법들은 연관규칙(Association Rule), 규칙기반필터링(Rules-based filtering), 협업필터링(Collaborative filtering), 학습에이전트(Learning agent) 등을 들 수 있다, 각각의 방법들은 실행하는 방법과 비용에서 차이를 가지고 있다. 본 연구에서는 데이터 마이닝의 알고리즘을 적용하여 개인화 추천 시스템을 제안한다.

2. 관련 연구

2.1 데이터 마이닝

데이터 마이닝은 대량의 데이터로부터 지식을 추출하는 것으로 쉽게 드러나지 않는 묵시적이고 잠재적인 정보까지 데이터베이스로부터 찾아내는 작업으로

+ 성창규 (한국해양대학교 대학원 컴퓨터공학과), E-mail:dolphin@bada.hhu.ac.kr, Tel: 051)410-4571

++ 류길수, 한국해양대학교 IT공학부

+++ 김태진, 한국해양대학교 대학원 컴퓨터공학과

서, 대용량의 데이터 내에 존재하는 데이터간의 관계, 패턴, 규칙 등을 탐색하고 찾아내어 모형화 함으로서 유용한 정보를 추출하는 과정이다.[10] 데이터 마이닝은 보다 더 정교하고 포괄적인 과정인 ‘지식 발견’ (KDD: Knowledge Discovery in Database) 과정의 일부이다.

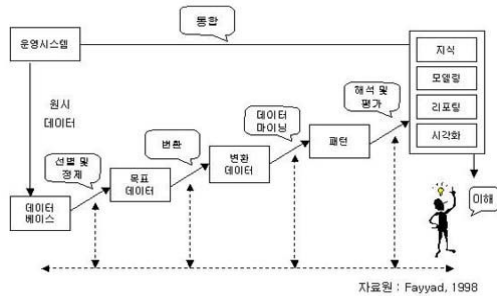


그림 1. KDD 프로세스

2.2 개인화 추천 방법

2.2.1 연관 규칙(Association Rule)

데이터 마이닝 기법 중 연관규칙 기법은 항목들의 빈도수와 동시 발생 확률을 이용하여 항목간의 관계를 찾아 연관성을 규칙으로 표현하는 기법이다. 연관규칙(규칙 R: IF A THEN B)에 대하여 지지도(Support(R))는 전체 트랜잭션 중 A와 B항목이 동시에 일어난 트랜잭션의 비율을 의미한다. 연관규칙의 강도를 나타내는 신뢰도(Confidence(R))는 조건 A를 만족하는 트랜잭션 중 B의 결론을 내릴 수 있는 트랜잭션의 비율을 의미한다.

$$\text{지지도} = \frac{A \cup B \text{를 포함하는 트랜잭션의 수}}{\text{트랜잭션 데이터베이스내 전체 트랜잭션 수}}$$

$$\text{신뢰도} = \frac{A \cap B \text{를 포함하는 트랜잭션의 수}}{A \text{를 포함하는 트랜잭션의 수}}$$

빈발항목집합은 일정수준 이상의 지지도를 가지는 모든 항목들의 집합을 의미하며 빈발항목집합이 추출되면 이를 이용하여 연관규칙을 생성한다. 연관 규칙이 최근 추천 시스템에 많이 적용될 뿐 아니라 데이터

의 연관성 찾아내는 연관 알고리즘에 대한 연구도 진행되었다.[1][8]

- 사용자 연관규칙

IF [사용자1:like] AND [사용자2:like]
THEN [목표사용자:like]
(신뢰도:90%, 지지도:70%)

→한 항목에 대하여 사용자1이 좋아하고, 사용자2가 좋아하면 목표 사용자도 좋아한다.

- 항목 연관규칙

IF [항목1:like] AND [항목2:like]
THEN [목표항목:like]
(신뢰도:90%, 지지도:60%)

→사용자가 항목1을 좋아하고, 항목2를 좋아하면 목표항목도 좋아한다.

그림 2. 추천을 위한 연관 규칙의 예

2.2.2 협업필터링(Collaborative filtering)

협업 필터링은 사용자가 자발적으로 제공한 정보를 사용하여 사용자를 비슷한 선호도를 가진 집단으로 나누어 그 집단 내에서 서로에게 추천하는 방식으로 얻어진 평가 데이터를 패턴으로 만들고 패턴 인식 기술을 사용해 서로 비슷한 선호도를 가진 집단으로 나눈다. 그룹에 속한 사람들끼리는 서로에 대해서 알지도 못하고 알 필요도 없다. 단지 나와 같은 그룹에 있는 선호도가 비슷한 사람이 내가 미처 보지 못한 어떤 것을 보고 그것이 좋다고 생각했다면 시스템은 그것을 추천한다.

2.2.3 규칙기반필터링(Rules-based filtering)

규칙기반 필터링은 사용자들에게 몇 가지 질문들을 한 이후에 이 질문들의 답에 적합한 내용들을 전달하는 것이다. 규칙기반 필터링에서 제공하는 질문은 사용자들이 자신의 선택을 통해 내용을 스스로 구성하게 하기 위한 목적이 아니라 질문을 통해 사용자들을 구분하고 개개인을 구별하기 위한 목적으로 사용된다.

2.2.4 학습에이전트(Learning agent)

학습 에이전트는 웹 사이트인 경우 사용자가 사이트 내에서 어떤 페이지에 오래 머무는지 등과 같은 사용자들의 행동을 기준으로 에이전트가 사용자의 선호도와 취향을 학습하여 알아내고 이를 바탕으로 사용자에게 적절한 상품을 추천한다.

3. 개인화 추천 시스템의 구현

3.1 개인화와 추천시스템

현재 개인화(personalization)라는 용어는 최근 인터넷의 대중화에 따라 협의의 의미로 웹사이트 개인화라는 의미로 많이 사용된다.[13]

추천 시스템(Recommend System)은 고객으로부터 직접 입수한 정보나 웹에서의 고객 행동 패턴, 구매 내역 등을 분석하여 획득한 정보를 바탕으로 고객이 차기에 구매할 가능성이 높은 상품의 리스트를 작성하여 고객에게 추천해 주는 시스템을 말한다.[2]

개인화 추천시스템의 활용분야로는 전자상거래 분야, 대규모 정보 제공 사이트, 포털 사이트, 회원제 운영 사이트 등의 많은 곳에서 활용이 늘어나는 추세이다.

3.2 개인화 시스템 구현

본 연구에서 제안한 시스템 구축환경은 구성은 아래 표와 같다.

구성 요소	세부 내용
운영체제	Microsoft Windows 2003
데이터베이스서버	Microsoft SQL Server 2005
분석 서버	Microsoft SQL Server 2005 Analysis Services
웹 서버	Microsoft IIS Server
사용 언어	ASP.NET

<표 1> 추천 시스템 구현 환경

데이터베이스 서버로는 Microsoft SQL Server 2005 Beta2를 사용하였다. 이 서버에 강력한 마이닝 기능을 지원해주므로 데이터베이스와 연동한 마이닝 처리가 훨씬 쉽게 처리할 수 있도록 지원해주고 있고 사용자 인터페이스는 ASP.NET으로 구축하였다.

3.2.1 사용 데이터

본 시스템에서 사용된 데이터는 관계형 데이터베이스의 구조로 되어 있으며 people, movie 등의 테이블들로 구성되어 있다. 데이터의 체계-관계도(ERD: Entity-Relationship Diagram)은 아래 그림과 같다.

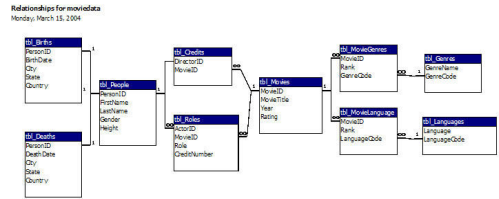


그림 3. 데이터베이스 체계-관계 모델

3.2.3 추천 시스템 개발

추천 시스템의 개발 절차는 다음과 같다.

- ① 데이터 마이닝을 위한 이전의 고객 데이터 준비
- ② 새로운 고객에게 추천을 위한 마이닝 모델을 설계
- ③ 분석 서버에 마이닝 모델을 사용하여 데이터 분석
- ④ 분석된 모델을 ASP.NET으로 연동하여 결과 출력

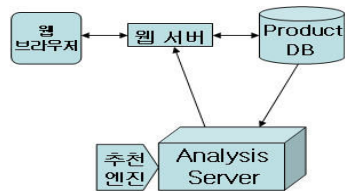


그림 4. 추천 시스템 구성

추천 시스템을 개발하기 위한 백-엔드 프레임워크(back-end framework)의 개발은 새로운 고객에게 추천을 위한 필요한 지식을 추출하는 것인데, 이 지식은 새로운 고객에게 추천을 위한 마이닝 모델에 내재되어 있다. 이 마이닝 모델은 과거 고객의 영화 구매 트랜잭션을 마이닝 처리함으로써 추출할 수 있다.

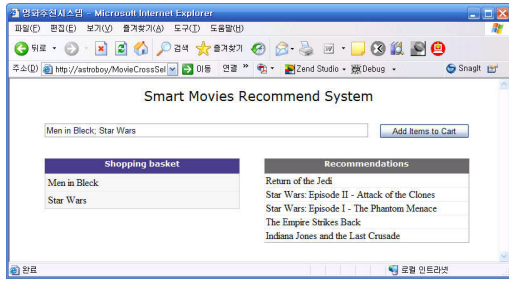


그림.5 영화 추천 시스템 실행화면

4. 결 론

본 연구에서는 데이터마이닝의 연관규칙 기법을 이용한 영화 추천시스템(MRS: Movies Recommend System)을 구현하였다. 여기서 eCRM 관련 연구 분야 중에서 고객에게 추천을 원활히 지원해줄 수 있는 방법을 제시하였다. 본 추천 시스템은 고객에게 추천할 대상 상품을 영화로 이용이 되었었지만 반드시 영화만으로 한정되는 것이 아니라 영화 이외의 다른 상품 추천에도 사용할 수 있으므로 적절한 여러 분야에 적용이 가능한 시스템이다. 개인화 추천 방법 중에서 각각의 방법들은 실행하는 방법과 비용에서 차이를 가지고 있다. 각 방법들의 비용과 직접/간접적인 효과를 정확하게 판단하고 적절한 방법을 선택하는 것이 매우 중요하다. 향후에 추천 시스템의 발전 과정에는 협업필터링(Collaborative filtering) 기법을 추가하고 고객 개인의 성향에 더 잘 맞도록 지능형 추천 시스템으로 발전시키는 것이다.

참고문헌

[1] R. Agrawal and R. Srikant, "Fast Algorithm for Mining Association Rules" Proc of 20th Int'l Conf. on VLDB, pp487-499, 1994

[2] Maltz D. and K. Ehrlich, "Pointing the Way:Active Collaborative Filtering," Proceedings of ACM CHI 1995 Conference on Human Factors in Computing System, 1996, 202~209

[3] Ansari, Asim, Skander Essegaier, and Rajeev Kohil, "Internet Recommendation Systems", Journal of Marketin Research, Vol.XXXVII, August(2000).

[4] Greg Linden, Brent Smith, Jeremy York "Amazon.com Recommendations" Published by the IEEE Computer Society Industry Report, January(2003)

[5] Rakesh Agrawal and Ramakrishnan Srikant, Mining generalized association rules" the VLDB Conference, Zurich, Switzerland, Septemper 1995

[6] Chumki Basuzy, Haym Hirshy, William WCohenx, Craig NevillManningy, "Recommending Papers by Mining the Web"

[7] Bamshad Mobasher, Honghua Dai, Tao Luo, Miki Nakagawa "Effective Personalization Based on Association Rule Discovery from Web Usage Data"

[8] Craig S., B. Sergey and M. Rajeev, "Beyond Market Baskets: Generalizing Association Rules to Dependence Rules", Data Mining and Knowledge Discovery, Vol.2 No.1, pp.39-68, 1998

[9] Han · Kamber "Data Mining Concepts and Techniques" 2000

[10] 김정자, 이도현, "데이터마이닝 기술 및 연구동향" 한국정보과학회 정보과학회지, 제16권, 제9호, 1998년 9월

[11] 박종수, 유원경, 홍기형, :연관규칙 탐사와 그 응용 " 한국정보과학회 정보과학지

제16권, 제9호 1998년 9월

- [12] 김재경, 안도훈, 서지혜, 조윤희 “협업 필터링 기법을 활용한 개인화된 상품 추천 방법론 개발에 관한 연구” 한국정보지능시스템학회 제8권 제2호 2002년 12월

- [13] 코리아 인터넷 마케팅센터,
"http://www.webpro.co.kr"

- [14] SQL Server Datamining 2005
http://www.sqlserverdatamining.com

- [15] 권병희, 김병직 공저 “about SQL Server
를 이용한 데이터 분석” 영진출판사

- [15] Ollie Cornes & Rob Birdwell 외, 성지용
외 옮김 “Beginning ASP.NET Using VB.NET
” 정보문화사