

웹 콘텐츠 추천 시스템 설계 및 구현

김산성^o, 류정우, 성지에, 차진호, 김명원
승실대학교 컴퓨터학부
kbibboss@hotmail.com, mkim@comp.ssu.ac.kr

Design and Implementation of a Web Contents Recommendation System

San-Sung Kim^o, Jung-Woo Ryu, Ji-Ae Sung, Jin-Ho Cha, Myoung-Won Kim
School of Computing, Soonsil University

요 약

디지털 콘텐츠 산업의 성장, 전자상거래의 활성화, 기업의 홈페이지 활용 증가 등으로 온.오프라인에서 콘텐츠의 수요가 증가하면서 콘텐츠를 관리하는 콘텐츠 관리 시스템 시장의 성장 잠재성이 높아지고 있다. 본 논문에서는 이러한 콘텐츠 관리 시스템의 마지막 단계인 콘텐츠 배포 단계에 있어 모든 사용자에게 동일한 콘텐츠를 제공하는 것이 아니라 사용자의 관심에 따라 다른 콘텐츠를 동적으로 제공하는 콘텐츠 추천 시스템을 설계 및 구현한다. 본 시스템은 규칙 기반 추천 방식을 사용하고 있으며 규칙으로는 사용자간의 연관성을 나타내는 사용자 협업적 규칙과 항목간의 연관성을 나타내는 항목 협업적 규칙이 존재한다. 또한 콘텐츠에 대한 사용자의 선호도를 측정하기 위해서 선호범위를 정의하고 있으며 취미, 관심분야와 같이 하나 이상의 값을 가질 수 있는 다중 값을 처리하고 있다. 시스템은 추천을 위한 정보 즉, 선호 범위와 사용자 프로파일 그리고 규칙들을 생성하는 오프라인 작업과 이러한 정보를 이용하여 실시간으로 사용자에게 추천해주는 온라인 작업으로 나뉘어 진다.

1. 서 론

디지털 콘텐츠 산업의 성장, 전자상거래의 활성화, 기업의 홈페이지 활용 증가 등으로 온.오프라인에서 콘텐츠의 수요가 증가하면서 콘텐츠를 관리하는 콘텐츠관리 시스템(Content Management System, 이하 CMS) 시장의 성장 잠재성이 높은 것으로 평가받고 있다. IDC에 따르면, 2000년도 세계 CMS 소프트웨어 시장규모는 약 20억달러로 전년 동기간에 비해 89.9%나 증가되었다고 보고하고 있다. CMS는 다양한 미디어 포맷에 따라 콘텐츠를 작성, 수집, 관리, 배포하는 콘텐츠 생산에서부터 활용, 폐기까지 즉, 콘텐츠 전 공급 과정에 관련된 애플리케이션이라 할 수 있다. 즉, 콘텐츠를 자체 제작하거나 외부로부터 수집하는 생성 단계에서는 자체 콘텐츠 작성물을 제공하거나 외부 업체의 콘텐츠 작성물을 지원하고, 업체 내외부의 콘텐츠를 통합하는 기능을 제공한다. 콘텐츠 관리단계에서는 워크플로우와 비즈니스 프로세스를 반영하여 콘텐츠를 관리하는 것뿐만 아니라 버전관리, 사용자권한관리, 편집, 콘텐츠 분류, 검색기능 등을 제공한다. 마지막 배포 단계에서는 사용자의 선호와 권한별로 개인화된 콘텐츠를 제공하는 단계이다.

본 논문에서는 CMS 마지막 단계인 콘텐츠 배포단계에 있어 모든 사용자에게 동일한 콘텐츠를 제공하는 것이 아니라 사용자의 관심에 따라 다른 콘텐츠를 동적으로 제공하는 콘텐츠 추천시스템을 설계한다.

본 시스템에서는 규칙 기반 추천방식을 사용하고 있으며, 규칙으로는 사용자간의 연관성을 나타내는 사용자 협업적 규칙(User Collaborative Rule)과 항목간의 연관

성을 나타내는 항목 협업적 규칙(Item Collaborative Rule) 두 종류가 존재한다. 또한 이러한 규칙들을 데이터베이스에서 생성하는 오프라인(Back-end) 작업과 생성된 규칙을 활용하여 사용자에게 콘텐츠를 실시간으로 추천하는 온라인(Front-end)작업으로 나누어진다.

본 논문의 구성은 다음과 같다. 2장에서 기존 추천시스템의 광범과 문제점을 살펴본다. 3장에서는 본 시스템의 특징에 대해 서술하고 4장에서 시스템의 구성을 살펴본다. 5장에서는 시스템의 실행 예를 살펴봄으로써 마지막으로 6장에서 결론 및 향후 연구에 대해 기술한다.

2. 관련연구

현 인터넷상에서 취향에 맞는 콘텐츠(상품 및 정보)를 사용자에게 추천해 주는 추천시스템은 사용하는 정보의 특성에 따라 내용기반 추천(contents-based recommendation), 인구 통계학적 추천(demographic recommendation), 협업적 추천(collaborative recommendation)으로 나눌 수 있다.[1] 내용기반 추천은 사용자에게 있어 이전에 선호한 항목과 비슷한 특성을 갖는 항목을 사용자 취향에 맞을 가능성이 높다고 보고 선호된 항목들의 속성 정보를 이용하여 추천하는 방법이며, 인구통계학적 추천은 사용자의 나이, 성별, 생활수준 등과 같은 인구통계학적 정보를 바탕으로 항목의 선호도를 추정하는 방법이다. 이에 비해 협업적 추천은 특정 사용자가 선호하는 항목들과 다른 사용자들이 선호하는 항목들을 바탕으로 항목의 선호도를 추정하는 방법이다.

초기 추천시스템은 협업적 추천 방법을 사용하여 추천을 하였으나 선호하는 항목이 적은 사용자 예를 들면 초기 사용자와 같은 경우에는 추천을 할 수 없거나 성능이 떨어지는 문제점을 가지고 있다. 이러한 문제점을 해결하기 위해 앞에서 서술한 세 가지 방법을 혼합하여 적용

1) 본 연구는 한국과학기술원 뇌신경정보학연구사업 및 아 이포스시스템즈(주)의 지원에 의하여 수행되었습니다.

하는 시스템이 개발되고 있다. 특히, 인구통계학적 추천과 내용기반 추천을 위해 규칙을 생성하여 추천하는 규칙기반 추천방식을 적용하고 있다. 하지만 대량의 정보에서 수동으로 규칙을 생성한다는 것은 많은 시간과 노력이 소요되며 또한 성능을 보장할 수 없는 문제점이 있다. 따라서, 본 시스템에는 자동 규칙 생성을 위해 데이터마이닝 기법 중 연관화 기법을 적용하고 있다. 이는 [2][3]에서 제안된 방법들을 응용하고 있으며 생성되는 규칙의 형태는 (그림.1)과 같다. 즉, [규칙1]은 사용자와 사용자간의 관련성을 [규칙2]은 콘텐츠와 콘텐츠간의 관련성을 나타내고 있으며, [규칙3]은 인구통계학적 정보와 콘텐츠간의 관련성을 [규칙4]은 콘텐츠 내용 정보와 사용자간의 관련성을 나타내고 있다. 본 시스템에서는 [규칙1]와 같은 형태를 사용자 협업적 규칙이라 정의하며, 그 외의 규칙들을 항목 협업적 규칙이라 정의한다.

<p>[규칙1] 만약 사용자A이고 사용자B이면 사용자C이다. : 사용자A와 사용자B가 선호하는 콘텐츠는 사용자C도 선호한다.</p> <p>[규칙2] 만약 콘텐츠A이고 콘텐츠B이면 콘텐츠C이다. : 콘텐츠A와 콘텐츠B를 선호하는 사용자는 콘텐츠C를 선호한다.</p> <p>[규칙3] 만약 성별이 남성이고 취미가 등산이고 취미가 스포츠이면 콘텐츠C이다. : 성별이 남성이고 취미가 등산이고 스포츠인 사용자는 콘텐츠C를 선호한다.</p> <p>[규칙4] 만약 키워드A가 전산이고 키워드가 후련이고 키워드가 학습이면 사용자C이다. : 키워드A가 전산이고 키워드가 후련이고 학습인 콘텐츠는 사용자C가 선호한다.</p>
--

그림 1. 생성되는 규칙의 형태

3. 시스템의 특징

본 시스템에서는 규칙 기반 추천 방식을 사용하고 있으며 추천 방법은 2장에서 기술한 세 가지 방법을 혼합하여 사용하고 있다.

특히, 실시간으로 사용자 취향에 따라 추천해 주기 위해 콘텐츠에 대한 사용자 선호도를 분석하는 모듈과 개인의 사용자에 대한 정보를 저장하는 사용자 프로파일을 설계하였으며, 속성 값이 집합인 다중 값(Multi Value)을 처리하고 있다.

3.1 선호도 분석

콘텐츠에 대한 선호도를 분석하기 위해 본 시스템에서는 사용자가 콘텐츠를 열람한 시간을 고려하여 선호도를 분석하고 있다. 즉, 콘텐츠를 열람한 시간이 관리자가 정의한 선호범위 안에 포함된다면 사용자는 그 콘텐츠를 선호한다고 가정한다. 이러한 선호범위를 정의하기 위해 (그림.2)와 같은 그래프를 제공하고 있다.

(그림.2)에서 x축은 체류시간(단위:초)을 의미하고 y축은 접속횟수를 의미한다. 즉, 접속횟수가 많은 체류시간 때가 콘텐츠를 선호하는 선호범위가 될 수 있다. 따라서 관리자는 접속횟수와 선호범위에 포함되는 트랜잭션의 비율을 참고로 최소 체류시간과 최대 체류시간을 결정하여 선호범위를 정의한다.

3.2 사용자 프로파일 [4]

사용자 개인의 정보를 저장하기 위해 본 시스템은 사용자 프로파일을 정의한다. 사용자 프로파일은 개인이 오랜 기간 동안 흥미를 갖는 정보(long-term interests)를

저장하는 장기기억(long-term memory)과 순간 흥미를 갖는 정보(short-term interests)를 저장하는 단기기억(short-term memory)으로 구분되어 있다. 장기기억에는 협업적 사용자 규칙이 저장되어 있고, 단기기억의 정보는 영구적으로 보존되는 것이 아니라 일정기간 보존되었다가 자동으로 사라지게 된다.

이러한 정보들은 생성된 규칙으로 추천할 때 사용되거나 후처리(postprocess). 예를 들면 사용자가 본 콘텐츠를 삭제하는 과정에 사용하게 된다



그림 2. 선호도 분석

3.3 다중 값(Multi Value) 처리

다중 값이란 속성 값이 집합인 것을 말한다. 예를 들면 취미, 관련분야, 콘텐츠 키워드와 같은 속성들을 의미한다. 즉, 사용자가 가질 수 있는 취미나 관심분야는 하나 이상 일 수 있고 또한 콘텐츠에 포함된 키워드 역시 하나 이상 일 수 있다. 본 시스템에서는 이와 같이 집합을 갖는 다중 값을 처리할 수 있도록 연관화 기법을 수정하였다. 그 결과 다중 값이 처리되어 생성되는 규칙의 형태는 (그림.1)에서 [규칙3], [규칙4]와 같다.

4. 시스템 구성

본 시스템은 추천을 위한 정보, 선호범위와 사용자 프로파일 그리고 지식베이스를 생성하는 오프라인 작업과 이러한 정보를 이용하여 실시간으로 사용자에게 추천해주는 온라인 작업으로 나누어진다.

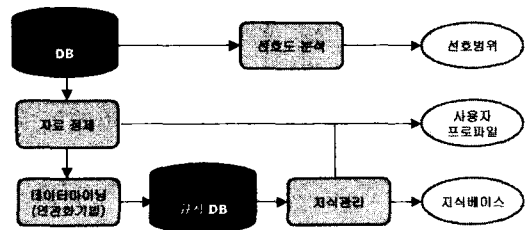


그림 3. 추천을 위한 정보 생성과정

오프라인 작업의 (그림.3)과 같다. 3장에서 서술한 선호범위를 트랜잭션 데이터를 이용하여 정의한다. 사용자가 관심 있게 본 콘텐츠 정보와 사용자 협업적 규칙들을 사용자 프로파일의 장기기억에 저장하고 자동으로 생성된 규칙들을 데이터베이스에 저장하고 지식관리모듈을 통해 관리자가 정제한 다음 지식베이스를 생성한다. (그림.4)

는 추천을 위한 정보 생성과정인 오프라인 작업을 위한 메인 화면을 보여주고 있다.

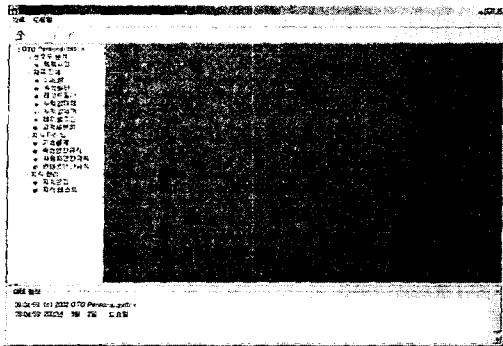


그림 4. 오프라인 작업을 위한 메인 화면

이와 같이 생성된 정보를 (그림5)와 같이 온라인 작업에 이용하여 사용자에게 실시간 추천이 이루어지도록 하고 있다. 사용자가 접속한 콘텐츠의 체류시간이 선호범위에 만족되면 사용자가 콘텐츠를 선호한다고 보고 콘텐츠에 대한 정보가 사용자 프로파일의 단기기억에 저장된다. 이러한 정보는 추천엔진에 사용자의 인구통계학적 정보와 같이 사실(fact)로 입력되어 추론이 이루어진다. 추론을 통해 생성된 추천 콘텐츠들은 후처리를 통해 사용자가 접속한 콘텐츠를 제거하거나 관심 있어하는 콘텐츠를 추가하여 최종 추천 콘텐츠 항목을 생성한다. 본 시스템의 추천엔진은 전문가 시스템 툴인 Jess6.0을 이용하고 있다.

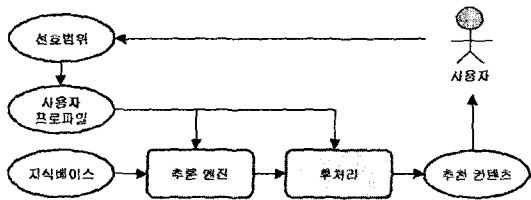


그림 5. 추천 콘텐츠 생성 과정

5. 시스템 실행 예

본 장에서는 시스템을 통해 추천되는 실행 예를 살펴본다. 추천을 위한 시나리오는 다음과 같다. "22세 남성인 사용자A는 대학생이며 미혼이고 취미가 여행이다. 사용자B는 23세의 미혼 여성이고 직장인이며 취미가 스포츠이다." (표.1)는 구축된 규칙들을 보여주고 있다. 여기서 [규칙7],[규칙8]은 지식관리를 통해 관리자가 생성한 규칙들이다.

사용자B가 접속하여 "[소식] 올해 카드 서비스 좋아진다" 콘텐츠를 관심 있게 보았다면 즉, 체류시간이 선호범위에 만족한다면 [규칙7]에 의해서 사용자B의 추천항목에 LG카드 콘텐츠가 추가되고 이때, 사용자A가 접속하면 [규칙1]에 의해 사용자A의 추천항목에 사용자B가 선호한 소식 콘텐츠가 추천되는 것을 확인할 수 있다. (그림.6)

표 1. 추천에 사용된 규칙

사용자 협업적 규칙	[규칙1] 만약 사용자B 이면 사용자A 이다.
항목 협업적 규칙	[규칙2] 만약 여성이고 20대이고 직장인 이면 결혼콘텐츠 이다
	[규칙3] 만약 남성이고 대학생 이면 영화 콘텐츠
	[규칙4] 만약 남성이고 여행을 좋아하며 대학생 이면 숙박 콘텐츠
	[규칙5] 만약 여성이고 스포츠를 좋아하면 스키콘텐츠
	[규칙6] 만약 결혼콘텐츠 이면 미용콘텐츠
	[규칙7] 만약 여성이고 소식콘텐츠 이면 LG카드 콘텐츠
	[규칙8] 만약 스키를 좋아하고 숙박 콘텐츠 이면 주유 콘텐츠

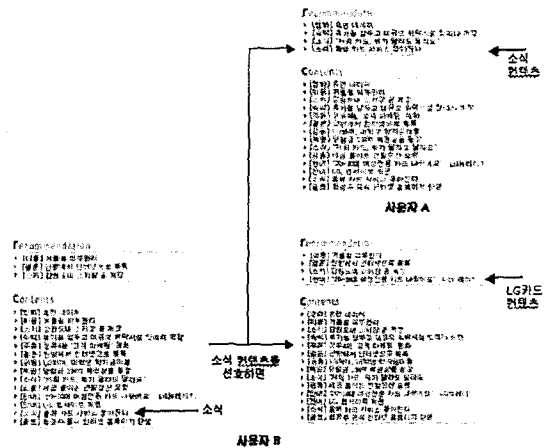


그림 6. 콘텐츠 추천 실행 예

6. 결론 및 향후연구

본 논문에서는 CMS의 마지막 단계인 콘텐츠 배포 단계에 있어 사용자의 취향이나 관심을 고려한 콘텐츠 추천 시스템을 설계 및 구현하였다. 본 시스템에서는 규칙 기반 추천 방식을 적용하였으며 대용량 데이터에서 추천을 위한 규칙을 자동으로 생성하기 위해 데이터마이닝 기법인 연관화 기법을 적용하였다. 또한 사용자 프로파일을 설계하여 개인 정보를 저장하였으며 콘텐츠에 대한 사용자 선호도를 측정하기 위해 선호범위를 정의하였다.

향후연구로 콘텐츠에 대한 선호도 분석을 위해 현재 시스템에서는 단지 접속횟수와 체류시간만을 고려하였으나 보다 정확한 분석을 위해 사용자 모니터링 에이전트를 설계하여 적용한다.

참고문헌

- [1] Michael J. Pazzani. "A Framework for Collaborative Content-Based and Demographic Filtering". *Artificial Intelligence Review* 13(5-6): pages 393-408, 1999
- [2] W. Lin, C. Ruiz, and S. A. Alvarez. "A new adaptive-support algorithm for association rule Mining". Technical Report WPI-CS-TR-00-13, Department of Computer Science, Worcester Polytechnic Institute, May 2000
- [3] Sarwar, B.M., Karypis, G., Konstan, J.A., and Riedl, J. "Item-based Collaborative Filtering Recommender Algorithms". Accepted for publication at the WWW10 Conference, May, 2001
- [4] D. Billsus, M.J. Pazzani, "A Hybrid User Model for News Story Classification", In Kay J.(ed.), Proceedings of the Seventh International Conference on User Modeling, Springer-Verlag, 99-108, 1999