

사례기반추론 기법을 이용한 개인화된 추천시스템 설계 및 구현

김 영 지[†] · 문 현 정^{††} · 옥 수 호^{†††} · 우 용 태^{††††}

요 약

본 논문에서는 인터넷 콘텐츠 사이트에서 묵시적인 평가정보를 이용한 새로운 사례기반 추천시스템을 설계하고 구현하였다. 본 시스템은 크게 사용자 프로파일 생성 모듈, 유사도 계산 및 추천 모듈, 개인화된 메일링 모듈로 구성된다. 사용자 프로파일 생성 모듈에서는 사용자가 콘텐츠를 이용하면서 남긴 로그 기록을 이용하여 콘텐츠에 대한 개인별 선호도를 추출할 수 있는 속성내, 속성간 가중치를 제시하였다. 유사도 계산 및 추천 모듈에서는 사용자 프로파일과 새로운 콘텐츠간의 유사도를 측정하기 위한 유사도 계산식을 제시하였다. 개인화된 메일링 모듈에서는 개인별 선호도에 의해 구성된 추천 콘텐츠를 플랫폼-독립적인 XML 문서 형식으로 변환하여 발송한다. 제안된 모델에 대한 추천 효율을 검증하기 위해 평균절대오차(MAE)와 반응자작용특성(ROC) 값을 이용하여 제안한 추천 모델과 협동적 필터링 기법과 비교 실험하였다. 실험 결과, 본 논문에서 제안한 모델의 추천 효율이 기존의 협동적 필터링 기법보다 우수함을 보였다.

Design and Implementation of personalized recommendation system using Case-based Reasoning Technique

Young-Ji Kim[†] · Hyeon-Jeong Mun^{††} · Soo-Ho Ok^{†††} · Yong-Tae Woo^{††††}

ABSTRACT

We design and implement a new case-based recommender system using implicit rating information for a digital content site. Our system consists of the User Profile Generation module, the Similarity Evaluation and Recommendation module, and the Personalized Mailing module. In the User Profile Generation Module, we define intra-attribute and inter-attribute weight derived from own's past interests of a user stored in the access logs to extract individual preferences for a content. A new similarity function is presented in the Similarity Evaluation and Recommendation Module to estimate similarities between new items set and the user profile. The Personalized Mailing Module sends individual recommended mails that are transformed into platform-independent XML document format to users. To verify the efficiency of our system, we have performed experimental comparisons between the proposed model and the collaborative filtering technique by mean absolute error (MAE) and receiver operating characteristic (ROC) values. The results show that the proposed model is more efficient than the traditional collaborative filtering technique.

키워드 : CRM(Customer Relationship Management), 추천시스템(Recommendation System), 사례기반추론(Case-based Reasoning), 개인화(Personalization)

1. 서 론

CRM 시스템에서 개인화(personalization) 전략은 개인별 로 맞춤 서비스를 제공하기 위한 핵심적인 전략중의 하나이다. 고객들에게 차별화된 서비스를 제공함으로써 고객들과

친밀한 관계를 유지하여 충성도(loyalty)를 높일 수 있다. 또한 콘텐츠 사이트에서 개인화 전략은 사용자들이 콘텐츠를 검색하는 시간과 경비를 절약하고 사이트에 대한 만족도를 높여 매출을 극대화하기 위한 전략으로 사용할 수 있다.

이러한 개인화 전략을 위한 핵심 기술중의 하나는 고객들의 취향과 구매 이력을 분석하여 개인별로 차별화된 정보를 자동적으로 필터링하기 위한 추천시스템이다. 추천시스템에 관한 기존의 연구 방법은 크게 인구통계학적 정보를 이용한 추천기법, 내용기반 추천기법, 항목기반 추천기법 그리고 가장 널리 사용되는 협동적 필터링(collaborative

* 본 논문은 정보통신부의 2001년 대학기초사업(2001-111-2) 지원으로 수행되었음.

† 준 회 원 : 창원대학교 대학원 컴퓨터공학과
 †† 준 회 원 : 창원대학교 대학원 전자계산학과
 ††† 정 회 원 : 고신대학교 전산학과 교수
 †††† 총신회원 : 창원대학교 컴퓨터공학과 교수
 논문접수 : 2002년 9월 30일, 심사완료 : 2002년 11월 21일

filtering) 기법 등이 있으며, 최근에는 여러 가지 형태의 추천 기법을 결합한 방법도 연구되고 있다.

하지만 기존의 추천 기법들은 다음과 같은 몇 가지 문제점을 가지고 있다. 첫째, 사용자들이 신상 정보의 유출을 염려하여 인적사항을 비워두거나 부정확한 내용을 입력하는 관계로 정확한 사용자 유형 분석이 어렵다. 둘째, 영화나 게임 등과 같이 개인별 선호도의 차이가 큰 콘텐츠에 대한 개인별 평가 정보를 온라인 상에서 수동적으로 수집하는 관계로 정확한 개인별 취향 분석이 어렵다. 셋째, 시간에 따라 변화하는 사용자의 개인별 선호도를 동적으로 반영할 수 있는 알고리즘 개발이 어렵다. 넷째, 디지털 콘텐츠의 종류가 다양해지고 대량으로 생산됨에 따라 개인별로 선호하는 콘텐츠를 사용자에게 제공하기 위한 자동화된 통합 솔루션이 부족하다.

본 논문에서는 사람이 과거에 경험한 사례를 이용하여 새로운 문제를 해결하는 사례기반추론(case-based reasoning) 기법을 이용한 새로운 형태의 추천기법을 제안하였다. 제안된 기법은 인터넷 콘텐츠 사이트에서 사용자에게 대한 최소한의 신상정보를 기반으로 콘텐츠에 대한 평가(rating) 정보를 자동적으로 수집하고, 콘텐츠에 대한 개인별 선호도를 동적으로 학습하여 새로운 콘텐츠를 추천하기 위한 기법이다. 그리고 사용자 편의를 위하여 개인별 선호도에 따라 추천된 콘텐츠를 플랫폼-독립적인 XML 문서 형식으로 변환하여 E-Mail로 발송하는 통합 시스템을 설계하고 구현하였다.

2. 관련 연구

2.1 기존 추천 기법의 연구 동향

추천기법에 관한 기존의 연구 동향은 크게 인구통계학적 정보에 기반한 추천 기법, 내용기반 추천 기법, 협동적 필터링 추천 기법 그리고 항목기반 추천기법 등으로 분류할 수 있다[1-3].

2.1.1 인구통계학적 추천(Demographic-based Recommendation) 기법

인구통계학적 정보에 의한 추천기법은 사용자의 성별, 나이, 직업 등과 같은 인구통계학적 요소에 의해 사용자 유형별 특징을 분석하여 상품을 추천하는 방법이다. 이 기법은 전통적인 추천기법의 하나로, 단순한 형태의 정보필터링 기법을 이용하는 타겟 마케팅 전략의 하나로 널리 사용되고 있다. 특히 이 기법은 사용자의 피드백 정보가 없이도 상품에 대한 추천이 가능하여 시스템 초기 구축 단계나 처음 방문한 사용자에게 대해서도 적용할 수 있다[3-5].

2.1.2 내용기반 추천(Content-based Recommendation) 기법

내용기반 추천기법은 개인의 요구나 개인으로부터 입력

된 모든 정보와 상품에 포함된 텍스트 정보를 이용하여 필터링하는 방식이다. 이 기법은 사용자 프로파일을 통해 과거 구매나 추천 결과를 쉽게 반영할 수 있는 장점이 있으며 추천 속도가 빠르다[6].

2.1.3 협동적 필터링(Collaborative Filtering) 기법

협동적 필터링 기법은 사용자의 상품에 대한 평가와 유사한 선호도를 가진 다른 사용자의 평가에 근거하여 사용자가 좋아할 만한 상품을 추천하는 기법이다. 이 기법은 숨어 있는 선호 패턴을 발견할 수 있으며 추천의 정확도가 높은 특징으로 인해 널리 사용되고 있다[7,8]. 협동적 필터링 기법에서 상품을 추천하는 과정은 크게 두 단계로 나뉘어진다. 먼저, correlation coefficient 식을 통해 사용자간의 유사도를 측정한다. 그리고 측정된 유사도를 이용하여 사용자별로 관심도를 예측한 후, 관심도가 높은 순으로 상품을 추천한다. 다음 식 (1)은 협동적 필터링 기법에서 유사도를 계산하는 식이며, 식 (2)는 관심도를 예측하는 식이다.

$$r_{KL} = \frac{Cov(K, L)}{\sigma_K \sigma_L} \quad (1)$$

$$K_{ipred} = \bar{K} + \frac{\sum_{J \in \text{neighbors}} (J_6 - \bar{J}) r_{KJ}}{\sum_J |r_{KJ}|} \quad (2)$$

2.1.4 항목기반 추천(Item-based Recommendation) 기법

항목기반 추천기법은 항목과 항목간의 관련성에 기반한 기법으로 사용자가 이미 구매한 항목을 근거로 그 항목과 관련이 있거나 유사한 항목을 추천하는 기법이다. 이 기법은 서로 다른 항목간의 유사도를 계산하기 위해 주로 cosine-based similarity와 conditional probability-based similarity를 이용한다. 이 기법은 각 항목에 대한 사용자의 선호도 값(1: 관심있음, 0: 관심없음)에 기반하여 user-item matrix를 생성한 후, cosine이나 probability 식을 이용하여 항목간의 유사도를 계산한다. 그리고 구매하지 않은 항목 중에서 이미 구매한 항목과 유사도가 높은 Top-N개의 항목을 추천한다. 아래 식 (3)과 식 (4)는 각각 cosine-based similarity와 conditional probability-based similarity를 계산하는 식이다[1].

$$\text{sim}(v, u) = \cos(\vec{v}, \vec{u}) = \frac{\vec{v} \cdot \vec{u}}{\|\vec{v}\|_2 \|\vec{u}\|_2} \quad (3)$$

$$\text{sim}(v, u) = \frac{\sum_{i: r_{i,v} > 0} r_{i,u}}{\text{Freq}(v) \times (\text{Freq}(u))^a} \quad (4)$$

2.2 기존 추천 기법의 문제점

그 동안 전자상거래나 콘텐츠 사이트에서 개인화된 상품

이나 콘텐츠를 추천하기 위한 다양한 추천 기법들에 대한 연구가 진행되고 있지만 다음과 같은 문제점을 가지고 있다. 먼저, 인구통계학적 정보에 의한 추천 기법은 인구통계학적인 정보에 의해 사용자 유형을 분류하기 위해 자세한 신상정보 수집 과정이 필요하다. 하지만 사용자들이 신상정보 유출을 꺼리는 경향으로 인해 사용자 유형 분석이 어렵다. 또한 사용자 유형에 따라 집단별 추천은 가능하지만 개인별 선호도를 반영한 개인화된 추천이 어렵다. 그리고 내용기반 추천 기법은 상품에 대한 텍스트 정보의 수집과 정확도에 대한 판단이 어렵고 상품과 사용자가 많은 경우에 추천 효율성이 저하되는 문제점이 있다[6].

그리고 협동적 필터링(collaborative Filtering) 기법은 초기 평가의 문제점, 희소성(sparsity), 동의어(synonym), 모호 집단(gray sheep) 등과 같은 문제점 등이 있다[9, 10]. 초기 평가의 문제점은 시스템 구축 초기에서 사용자로부터 충분한 평가 정보를 받지 못한 경우 정확한 추천이 어려운 문제이다. 희소성의 문제는 초기 시스템에서 발생하는 문제와 유사하며, 상품에 대해서 충분한 평가 결과를 구성하기 어려운 관계로 사용자와 평가 정보로 구성된 행렬이 희소성을 가지는 문제이다. 동의어 문제는 상품명이나 다르지만 유사한 상품에 대해 숨어있는 연관관계를 찾기 어려운 경우이다. 모호 집단의 문제는 좋은 것과 싫은 것이 분명하지 않거나 평가 결과가 일정하지 않은 특이한 사용자들에 대해서는 유사 집단의 발견이 어려운 관계로 추천이 어려운 문제이다[6].

2.3 사례기반추론 기법

인공지능분야에서 주로 사용되는 사례기반추론 기법은 사람이 새로운 문제를 해결하기 위해 과거에 경험한 사례

중에서 가장 유사한 사례를 검색하여 새로운 문제에 대한 해결과 교정을 통한 학습을 반복하는 기법이다. 이 기법은 새로운 문제가 발생하면 이미 경험한 사례 데이터베이스에서 가장 유사한 사례를 검색하여 이전 사례를 재사용하는 적용 과정을 통해 새로운 문제를 해결하는 방식이다. 또한 교정과정에서 새로운 문제의 해가 문제 해결에 적합한지를 검증하여 새로운 사례로 학습한다[11]. (그림 1)은 일반적인 사례기반추론 기법의 개념도이다.

사례기반추론 기법에서 가장 유사한 사례를 찾는 방법으로 Nearest-Neighbor기법과 귀납적 기법을 주로 사용한다. 아래 식 (5)은 Nearest-Neighbor기법에서 입력문제 T 와 사례 S 에 대한 유사도를 계산하는 식이다. 식 (5)에서 T 는 입력문제, S 는 학습된 사례, W_i 는 T 와 S 의 각 속성에 대한 가중치로 정의된다.

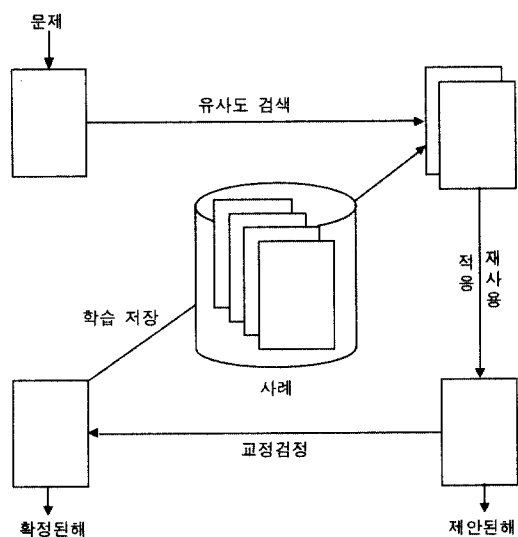
$$\text{Similarity}(T, S) = \sum_{i=1}^n f(T_i, S_i) \times W_i \quad (5)$$

3. 사례기반추론 기법을 이용한 추천시스템

본 논문에서는 사례기반추론 기법을 이용하여 인터넷 콘텐츠 사이트에서 콘텐츠 유형에 대한 개인별 선호도를 학습하여 새로운 콘텐츠를 개인 메일로 추천하는 시스템을 제안하였다. 이러한 사례기반추론 기법은 협동적 필터링과 달리 유사집단의 평가정보를 이용하지 않고, 개인별 속성에 대한 가중치와 속성 값을 이용하여 정보를 추천할 수 있다. 또한 이 기법은 사용자가 선호하는 콘텐츠 유형에 대한 속성을 추천에 반영할 수 있는 장점이 있으며, 사용자 프로파일 이용을 이용하여 개인화된 추천이 가능하다.

제안 시스템에 대한 아이디어는 다음과 같다. 먼저, 본 시스템에서는 사용자에게 대한 최소한의 신상정보와 콘텐츠를 구성하는 속성 정보에 대한 개인별 로그 기록을 이용하여 콘텐츠 유형에 대한 개인별 선호도를 추출할 수 있다. 콘텐츠 유형에 대한 개인별 선호도를 추출하기 위하여 속성내 가중치(Intra-attribute Weight)와 속성간 가중치(Inter-attribute Weight)를 제시하였다. 속성내 가중치와 속성간 가중치는 개인이 이용하는 콘텐츠 유형에 대한 선호도를 의미한다. 이 값은 개인이 콘텐츠를 이용하는 동안 남긴 로그 기록을 이용하여 계산되어 프로파일 데이터베이스에 저장된다. 또한 새로운 콘텐츠에 대한 개인화된 추천을 위하여 프로파일 데이터베이스에 저장된 콘텐츠 유형별 선호도에 대한 유사도를 계산하는 함수를 제시하였다. 그리고 추천 효율 향상과 사용자 편의성을 위하여 개인별 선호도에 따라 구성된 콘텐츠를 플랫폼-독립적인 XML 문서 형식으로 변환하여 E-Mail에 의해 발송할 수 있는 통합 시스템을 구현하였다.

특히 본 시스템은 새로운 사례에 대한 학습을 통하여 시



(그림 1) 사례기반추론 기법의 개념도

간에 따라 개인별 선호도의 변화를 동적으로 반영할 수 있다. 그리고 사용자가 콘텐츠를 이용하는 동안 남긴 로그 기록으로부터 평가정보를 묵시적으로 수집하는 관계로 사용자의 불편을 해소할 수 있다. 제안 시스템은 크게 사용자 프로파일 생성 모듈, 유사도 계산 및 추천 모듈, 개인화된 메일링 모듈로 구성된다. 다음 (그림 2)는 본 논문에서 제안한 추천 시스템의 개념도이다.

(그림 2) 제안 시스템의 개념도

3.1 사용자 프로파일 생성 모듈

본 모듈은 임의의 사용자가 콘텐츠를 이용하는 동안 남긴 로그 기록을 이용하여 콘텐츠에 대한 선호도를 계산하여 사용자 프로파일을 구성하는 기능을 수행한다. 사용자 프로파일은 $P = \{u, A, W, D\}$ 로 정의되며 u 는 사용자 아이디, A 는 콘텐츠 속성의 집합, W 는 속성내의 요소 값 각각의 가중치의 집합, D 는 속성간의 가중치의 집합을 의미한다.

3.1.1 속성내 가중치(Intra-attribute Weight)

임의의 콘텐츠가 n 개의 속성 A_1, A_2, \dots, A_n 을 가지고, 속성 A_i 는 m 개의 서로 다른 속성 값 $a_{i1}, a_{i2}, \dots, a_{im}$ 을 가질 때, 속성 A_i 에서의 속성내 가중치 W_i 는 $W_i = \{w_{i1}, w_{i2}, \dots, w_{im}\}$ 로 정의하였다. 이 때 각 속성 값 a_{ij} 에 대응하는 속성내 가중치 w_{ij} 는 다음 식 (6)과 같다.

$$w_{ij} = \frac{k_{ij}}{\sum_{p=1}^m k_{ip}} \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, m \quad (6)$$

이 때, k_{ij} 는 속성 값 a_{ij} 를 액세스한 횟수를 나타낸다. w_{ij} 는 사용자가 열람한 전체 콘텐츠 중에서 각 속성 A_i 의 속성 값 a_{ij} 가 출현한 빈도율을 계산한 값이다. 즉, 속성 A_i 내에서 특정 속성 값 a_{ij} 값의 속성내 가중치 (w_{ij})가 가

장 높으면 이 속성 값 a_{ij} 는 속성 A_i 중에서 가장 중요하지 않다.

3.1.2 속성간 가중치(Inter-attribute Weight)

속성간 가중치는 속성 집합 A 에 대한 통계적 상관관계 집합 $D = \{d_1, d_2, \dots, d_n\}$ 이다. 속성 A_i 의 속성간 가중치 d_i 는 다음 식 (7)와 같다.

$$d_i = \left| y_i - \frac{1}{m} \sum_{j=1}^m w_{ij} \right|, \quad y_i = \max_{1 \leq j \leq m} (w_{ij}), \quad i = 1, 2, \dots, n \quad (7)$$

A_i 의 속성간 가중치가 높으면 사용자는 모든 속성 중에서 A_i 속성을 가장 중요하게 생각한다고 해석할 수 있다. 속성 A_i 의 속성간 가중치가 0이면, 이 속성은 현 사용자에게 무의미한 속성임을 의미한다.

예를 들어, 채용정보 사이트에서 임의의 사용자가 직업 정보를 열람할 경우, 사용자들이 직업을 선택하는 주요 속성이 전공분야, 채용유형, 근무지 속성이라 가정한다. 만일 Smith라는 사용자가 채용정보 콘텐츠를 10번 열람하는 동안 전공분야에서는 DB분야 10회, 인공지능 3회, 통신 2회를 열람하였고, 채용유형에서는 교수 6회, 연구원 6회, Post-Doc 6회, 그리고 근무지에서는 서울 8회, 대구 2회가 열람하였다고 가정한다. 이 경우 속성내 가중치와 속성간 가중치 값에 의해 Smith는 콘텐츠 속성 중에서 근무지나 채용유형보다는 전공분야 속성에 가장 관심을 가지고 있는 것으로 분석할 수 있다. 그리고 채용유형에 대해서는 모든 경우에서 동일한 관심도를 보이는 관계로 무의미한 속성임을 알 수 있다.

위 예에서는 특정 분야의 콘텐츠에 대한 대표적인 속성에 대해 속성내 가중치와 속성간 가중치를 적용하였지만, 식 (6)과 식 (7)에서처럼 가중치 계산을 위해 적용할 수 있는 속성은 응용분야에 따라 임의의 개수로 확장 가능하다.

<표 1> 사용자 프로파일 구성 예

<표 1>은 Smith에 대해 속성내 가중치, 속성간 가중치 값을 포함하는 사용자 프로파일의 예이다.

이러한 방식으로 콘텐츠 사이트에서 모든 사용자들이 관심 있는 콘텐츠를 이용하는 동안 남긴 로그 기록을 이용하여 콘텐츠 유형에 대한 개인별 선호도를 자동적으로 계산할 수 있다. 콘텐츠 유형별로 계산된 개인별 선호도는 사용자 프로파일 데이터베이스에 저장된다.

3.2 유사도 계산 및 추천 모듈

본 논문에서 제시한 추천 기법은 다른 사용자의 평가에 의존하는 협동적 필터링 기법과 달리, 사용자 개인별 행동 패턴을 분석하여 추천을 하기 위한 기법의 하나로 사례기반추론 기법과 가장 유사한 모델이다. 이에 따라 본 논문에서는 사례기반 추론 기법의 하나인 Nearest-Neighbor기법에서 사용하는 유사도 평가 방법을 확장하여 임의의 속성 값에 대해 속성내 가중치와 속성간 가중치를 이용한 유사도 평가 방법을 새롭게 제시하였다. 본 방법은 개인별 행동 패턴을 자세하게 분석할 수 있는 장점을 가진다.

본 모듈은 임의의 사용자에게 대하여 새로운 콘텐츠를 추천하는 기능을 수행한다. 즉, 새로운 콘텐츠와 사용자 프로파일에 저장된 콘텐츠 유형별 선호도간의 유사도를 계산하여 가장 높은 순으로 콘텐츠를 추천한다.

먼저, 새로운 콘텐츠 집합 I 는 $I = \{A'\}$ 로 정의한다. A' 은 A_i 의 집합이며 A_i 은 k 개의 속성을 가지는 $A'_i = \{a'_{i1}, a'_{i2}, \dots, a'_{ik}\}$ 로 표현된다. 이 때, 사용자 프로파일의 속성 A 의 도메인과 새로운 콘텐츠 집합의 속성 A'_i 의 도메인은 서로 동일하다. 다음 식 (8)은 사용자 프로파일 P 와 새로운 콘텐츠 집합 I 사이의 유사도를 계산하는 식이다.

$$\text{Similarity}(P, I) = \sum_{i=1}^n (f(a_{ij}, a'_{ij}) \times d_i \times w_{ij}) \quad (8)$$

(그림 3) 신규 콘텐츠

여기서, $f(a_{ij}, a'_{ij})$ 는 사용자 프로파일의 속성 값과 새로운 콘텐츠의 속성 값을 비교하는 함수로, 새로운 콘텐츠의 속성 값이 사용자 프로파일내에 있으면 1의 값을 반환하고, 없으면 0을 반환한다. 따라서, $f(a_{ij}, a'_{ij})$ 값이 1인 경우에만 해당 속성 값의 속성내 가중치와 속성간 가중치를 이용하여 콘텐츠에 대한 가중치를 계산한다.

다음 (그림 3)는 새로운 콘텐츠 채용정보에 대한 예이다.

<표 1>의 Smith의 사용자 프로파일 정보와 (그림 3)의 신규콘텐츠에 대해 식 (8)에 의해 유사도를 계산하면 채용정보 #101은 0.1, 채용정보 #102는 0.84, 채용정보 #103은 0.34가 된다. 따라서 본 시스템은 Smith에게 #102, #103, #101 순으로 새로운 콘텐츠를 추천한다.

3.3 개인화된 메일링 시스템 모듈

본 모듈은 추천 효율의 향상과 사용자 편의를 위해 개인별 선호도에 따라 구성된 추천 콘텐츠 목록을 주기적으로 메일을 발송하는 기능을 수행한다. 또한 본 모듈은 추천 콘텐츠를 다양한 형태의 콘텐츠로 구성할 수 있도록 플랫폼에 독립적인 XML 기반으로 개발하여 PDA, 핸드폰 등에도 적용할 수 있다. 다음 (그림 4)는 개인화된 메일링 시스템 모듈의 개념도이다.

(그림 4) 개인화된 메일링 시스템 모듈의 개념도

개인별 선호도에 따라 구성된 추천 콘텐츠를 메일로 발송하는 과정은 다음과 같다.

3.3.1 메일 발송 스케줄링

개인별로 주기적으로 메일을 발송하기 위한 스케줄링을 생성하는 단계이다. 본 시스템에서는 오라클의 Job Queue 기능을 사용하여 메일 발송 스케줄링을 자동화하는 방법을 개발하였다.

3.3.2 개인별 추천 콘텐츠 발송용 메일 작성

추천 시스템에 의해 생성된 개인별 추천 콘텐츠 목록 데이터베이스와 사용자 데이터베이스에 의해 XML 문서를 생성한

다. XML 문서는 XSLT를 통해 XSL에 정의된 Style Sheet에 따라 HTML 문서로 변환된다.

3.3.3 개인별 메일 발송

개인화된 추천 콘텐츠 목록에 대한 HTML 문서를 SMTP 서버를 통해 개인 메일로 발송한다.

4. 실험 및 고찰

본 연구에서 제안한 사례기반추론 추천 기법의 효율성을 검증하기 위하여 채용정보 사이트(http://www.hibrain.net)에서 운영중인 채용정보 콘텐츠에 대해 실험하였다. 먼저, 824 명의 사용자가 총 1,144건의 채용정보 콘텐츠에 대해 평균 10건 정도씩 열람한 웹 로그를 분석하여 사용자 프로파일 데이터베이스를 작성하였다. 그리고 추천 효율을 측정하기 위해 1,484건의 새로운 콘텐츠를 사용하였다.

추천 효율을 측정하는 방법은 정확도를 기준으로 MAE (Mean Absolute Error : 평균절대오차)와 ROC(Receiver Operating Characteristic : 반응자작용특성) 측정법이 있다[12]. MAE는 통계적인 정확도를 측정하는 방법으로 아이템에 대한 사용자의 평가 정보와 수치화된 예측 값을 비교하기 위해 사용되며, 예측된 평가 값의 MAE는 최소화되어야 정확도가 높다고 할 수 있다. 대상집합의 실제 평가 값을 $\{r_1, \dots, r_n\}$ 이라고 한다면, 동일한 아이টে에 대한 예측 값은 $\{p_1, CDTS, p_n\}$ 으로 표현하고, 오차 $E = \{e_1, \dots, e_n\} = \{p_1 - r_1, \dots, p_n - r_n\}$ 이라면, MAE는 아래와 같은 식으로 정의한다.

$$| \overline{E} | = \frac{\sum_{i=1}^n |e_i|}{n} \quad (9)$$

ROC는 Decision-support의 정확도를 측정하는 방법의 하나로 사용자가 가장 양질의 아이টে를 선택할 수 있도록 어떻게 효율적으로 예측하는지 여부를 평가하는 방법이다. ROC는 Sensitivity, Specificity, Accuracy, Error rate에 의해 평가할 수 있다. ROC를 사용할 경우 정확도 분석을 위해 good과 bad 콘텐츠를 구분하는데 사용자의 rating 정보 중 4, 5의 값을 good(positive), 1, 2, 3의 값을 bad(negative)라고 정의한다. True는 예측 값이 올바르게 평가된 것을 의

미하고, false는 예측 값이 잘못 평가된 것을 의미한다[12].

Sensitivity는 임의로 선택된 평가 값이 4이상인 콘텐츠가 추천될 확률이고 다음 식과 같이 정의된다.

$$Sensitivity = \text{true positive} / \text{total actual positive} \quad (10)$$

Specificity는 임의로 선택된 평가 값이 3이하인 콘텐츠가 추천되지 않을 확률이고 다음 식과 같이 정의된다.

$$Specificity = \text{false negative} / \text{total actual negative} \quad (11)$$

Accuracy는 추천의 정확도를 의미하며 다음 식과 같이 정의된다.

$$Accuracy = (\text{true negative} + \text{true positive}) / \text{grand total} \quad (12)$$

Error rate는 추천의 오차율을 의미하여 다음 식과 같이 정의된다.

$$Error\ rate = (\text{false negative} + \text{false positive}) / \text{grand total} \quad (13)$$

다음 <표 2>는 본 연구에서 제안한 사례기반추론을 이용한 추천 기법과 기존의 협동적 필터링 기반의 추천 시스템과 MAE 및 ROC 측정값을 비교한 결과이다. 실험 결과에서처럼 본 논문에서 제안한 시스템의 추천 효율이 협동적 필터링 기법보다 높은 정확도를 보였다.

<표 2> 제안 시스템과 협동적 필터링 기법간의 정확도 비교

구 분	MAE	ROC			
		Sensitivity	Specificity	Accuracy	Error rate
협동적필터링 기법	.923	.667	.210	.701	.299
제안 시스템	.818	.900	.424	.809	.191

다음 <표 3>은 제안 시스템과 기존의 다양한 추천 기법간의 장단점을 비교 분석한 내용이다.

5. 결 론

본 논문에서는 인공지능 분야에서 주로 사용되는 사례기반추론 기법을 이용하여 인터넷 콘텐츠 사이트에서 콘텐츠

<표 3> 제안 시스템과 기존 추천 기법간의 장단점 비교

구 분	개인화 추천	평가 정보	추천 방식	동적 학습 기능	추천 효율
인구통계학적 기법	불가능	불필요	신상정보를 이용한 집단 추천	어려움	낮다
콘텐츠 기반 추천	어려움	텍스트 정보	상품에 대한 부가정보 이용	어려움	대량데이터에서 효율 저하
협동적필터링 기법	가능	필요	타인의 평가정보 이용한 상관 계수	어려움	평가정보가 충분할 경우에 효율적
제안 시스템	가능	자동 수집	본인선호도를 이용한 지능형 추천	가능	개인별 평가정보가 충분할 경우에 효율적

유형에 대한 개인별 선호도를 학습하여 새로운 콘텐츠를 개인 메일로 추천하기 위한 시스템을 구현하였다. 본 시스템은 사용자에 대한 최소한의 신상정보와 콘텐츠를 구성하는 속성 정보에 대한 개인별 접속기록을 이용하여 콘텐츠 유형에 대한 개인별 선호도를 학습할 수 있다. 속성내 가중치(Intra-attribute Weight)와 속성간 가중치(Inter-attribute Weight)를 정의하여 개인이 콘텐츠를 이용하는 동안 남긴 로그 기록을 이용하여 콘텐츠 유형에 대한 개인별 선호도를 추출하여 사례 데이터베이스로 활용하였다. 그리고 새로운 콘텐츠와 프로파일 데이터베이스에 저장된 콘텐츠 유형별 선호도에 대한 유사도를 계산하는 함수를 제시하여 개인별 추천 콘텐츠 목록을 구성하였다. 그리고 추천 효율 향상과 사용자 편의성을 위하여 개인별 선호도에 따라 구성된 추천 콘텐츠를 플랫폼-독립적인 XML 문서 형식으로 변환하여 E-Mail에 의해 발송할 수 있는 통합 시스템을 구현하였다. 본 연구에서 제안한 추천 기법과 기존의 협동적 필터링 기반의 추천 시스템과 MAE 및 ROC 측정 값에 대한 비교 실험을 통하여 제안한 시스템의 추천 효율이 협동적 필터링 기법보다 우수함을 입증하였다. 특히, 본 시스템은 다른 사용자의 평가 정보를 이용하는 협동적 필터링 기법과는 다르게 콘텐츠 유형에 대한 자신의 선호 패턴을 이용하여 추천이 가능하다. 또한 새로운 사례에 대한 학습을 통하여 시간에 따라 개인별 선호도의 변화를 동적으로 반영할 수 있다. 그리고 사용자가 콘텐츠를 이용하는 동안 평가정보를 묵시적으로 수집하여 평가정보 수집에 따른 사용자의 불편을 해소할 수 있다.

앞으로 본 시스템은 고객과의 지속적인 관계를 유지하기 위한 개인화 전략, 추천효율 분석을 통한 일대일 마케팅 분야에 응용할 수 있을 것으로 기대된다. 또한 지식정보화 사회에서 국가경쟁력의 원천으로 인식되고 있는 온라인 게임, 교육용 콘텐츠, 인터넷 영화, 인터넷 음악, e-book 등과 같은 응용분야에서 활용 가능한 실용적인 시스템 개발과 콘텐츠 사이트의 수익 모델창출에 기여할 수 있으리라 생각한다.

참 고 문 헌

[1] G. Karypis, "Evaluation of Item-Based Top-N Recommendation Algorithms," Technical Report CS-TR-00-46, Computer Science Dept., University of Minnesota, 2000.
 [2] B. M. Sarwar, G. Karypis, J. A. Konstan and J. Riedl, "Analysis of Recommender Algorithms for E-Commerce," Proc. of the 2nd ACM E-Commerce Conference, pp.158-167, 2000.
 [3] 한정기, "개인화(Personalization)의 핵심 기술", [http://personalization.co.kr/column\[010319\].htm](http://personalization.co.kr/column[010319].htm), 2001.
 [4] B. Krulwich, "LIFESTYLE FINDER : Intelligent User Profiling Using Large-Scale Demographic Data," Artificial

Intelligence Magazine Vol.18, No.2, pp.37-45, 1997.
 [5] M. J. Pazzani, "A Framework for Collaborative, Content-Based and Demographic Filtering," Artificial Intelligent Review, pp.394-408, 1999.
 [6] M. Claypool, A. Gokhale, T. Miranda, P. Murnikov, D. Netes and M. Sarti, "Combining Content-Based and Collaborative Filters in an Online Newspaper," ACM SIGIR Workshop on Recommender Systems, Berkeley, CA, 1999.
 [7] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based Collaborative Filtering Recommendation Algorithms," Proc. of the Tenth International World Wide Web Conference on World Wide Web, pp.285-295, 2001.
 [8] Rachael Rafter, Keith Bradley, Barry Smyth, "Personalised Retrieval for Online Recruitment Services," In Proceedings of the 22nd Annual Colloquium on Information Retrieval (BCS-IRSG 2000), Sidney Sussex College, Cambridge, UK, 2000.
 [9] Mark Claypool, David Brown, Phong Le and Makoto Waseda, "Inferring User Interest," Technical Report WPI-CS-TR-01-97, May, 2001.
 [10] J. B. Schafer, J. Konstan and J. Riedl, "Recommender Systems in E-Commerce," ACM Conference on Electronic Commerce, pp.158-166, 1999.
 [11] R. Schank, "Dynamic Memory : A Theory of Learning in Computers and People," Cambridge University Press, New York, 1982.
 [12] 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, pp.439-446, 1999.

김 영 지

e-mail : yjkim@cdcs.changwon.ac.kr
 1997년 창원대학교 전자계산학과 졸업 (이학사)
 1999년 창원대학교 전자계산학과 졸업 (이학석사)
 2001년~현재 창원대학교 컴퓨터공학과 박사과정 재학중

관심분야 : 데이터마이닝, 검색엔진, 추천 알고리즘, e-Learning

문 현 정

e-mail : mun@sarim.changwon.ac.kr
 1994년 한국방송대학교 전자계산학과 졸업 (이학사)
 1996년 창원대학교 전자계산학과 졸업 (이학석사)
 2001년 창원대학교 전자계산학과 박사수료

관심분야 : KDD, 텍스트마이닝, e-Learning

옥 수 호

e-mail : shok@kosis.ac.kr

1982년 경북대학교 전자공학과 졸업(공학사)

1984년 경북대학교 전자공학과 졸업(공학석사)

1998년 경북대학교 컴퓨터공학과 졸업(공학박사)

1988년~2000년 동의공업대학 부교수

2000년~현재 고신대학교 전산학과 조교수

관심분야 : Temporal Database, 데이터웨어하우스, 추천 알고리즘

우 용 태

e-mail : ytwoo@sarim.changwon.ac.kr

1982년 경북대학교 전자공학과 졸업(공학사)

1984년 경북대학교 전자공학과 졸업(공학석사)

1995년 경북대학교 전자공학과 졸업(공학박사)

1987년~현재 창원대학교 컴퓨터공학과 정교수

관심분야 : 데이터마이닝, CRM, 추천 알고리즘, e-Learning