

# 목시적 평가정보를 이용한 사례기반추론 추천시스템

김병찬<sup>0</sup>, 옥수호<sup>2</sup>, 우용태<sup>1</sup>

<sup>1</sup> 창원대학교 컴퓨터공학과, <sup>2</sup> 고신대학교 전산수학과

bckim@cdcs.changwon.ac.kr, shok@kosin.ac.kr, ytwoo@sarim.changwon.ac.kr

## A Recommender System using Case-based Reasoning with Implicit Rating Information

Byeong-Chan Kim<sup>0</sup>, Sooho Ok<sup>2</sup>, Yong-Tae Woo<sup>1</sup>

<sup>1</sup> Dept. of Computer Engineering, Changwon National University

<sup>2</sup> Dept. of Computer Science, Kosin University

### 요 약

본 논문에서는 인터넷 콘텐츠 사이트에서 개인별로 콘텐츠를 효과적으로 추천하기 위한 개인화 시스템 모델을 제안하였다. 제안한 모델은 목시적인 평가정보를 이용한 사례기반추론 기법으로서 협동적필터링 기법과 달리 유사집단의 평가정보를 이용하지 않고 개인별 속성에 대한 가중치와 속성 값을 이용하여 추천하는 기법이다. 이 기법은 각 사용자의 상품 구매 속성을 추천에 반영할 수 있는 장점이 있으며, 사용자 프로파일을 이용하여 개인화된 추천이 가능하다. 제안한 기법이 Recall, Precision, F-measure의 평가 방법을 통해 실험한 결과 협동적필터링 기법 보다 모든 부분에서 더 좋은 결과가 나왔음을 볼 수 있다. 그러므로 제안 시스템이 유사 사용자의 평가정보를 이용한 협동적필터링 기법보다 효율적인 개인화 전략이 가능하다고 할 수 있다. 본 제안 모델을 이용하여 일대일 마케팅을 위한 eCRM 시스템 개발이 가능하리라 예상된다.

### 1. 서 론

인터넷 사이트에서 개인화된 서비스를 제공하기 위한 핵심적인 요소 중의 하나는 고객들의 구미에 맞는 상품이나 콘텐츠를 개인별로 추천하기 위한 시스템이다. 추천시스템은 고객들의 상품에 대한 선호도와 구매이력을 분석하여 개인별로 차별화된 정보를 추천하기 위한 자동화된 정보필터링 시스템이다.

이러한 추천시스템 중에서 사례기반추론 기법은 유사집단의 평가정보를 이용하지 않고 개인별 속성에 대한 가중치와 속성 값을 이용하여 추천하는 기법이다. 이 기법은 각 사용자의 상품 구매 속성을 추천에 반영할 수 있는 장점이 있으며, 사용자 프로파일을 이용하여 개인화된 추천이 가능하다[1, 2].

사례기반추론 기법에서 사용자 프로파일은 상품이나 콘텐츠에 대한 개인별 선호도를 의미한다. 이러한 사용자 프로파일은 사용자로부터 명시적인 방법에 의해 수집하거나 사용자 별로 사이트 접속 패턴을 이용하여 목시적인 방법으로 수집할 수 있다[3]. 명시적인 정보를 이용한 추천은 사용자의 평가정보가 부족하거나 평가기준이 바뀔 경우 추천 효율이 낮아질 수 있다. 하지만 목시적인 평가정보는 사용자의 행동패턴에 따라 평가정보를

자동적으로 수집하는 관계로 명시적인 방법의 단점을 보완 할 수 있다.

본 논문에서는 인터넷 콘텐츠 사이트에서 개인별로 콘텐츠를 효과적으로 추천하기 위한 개인화 시스템 모델을 제안하였다. 제안한 모델은 목시적인 평가정보를 이용한 사례기반추론 기법을 기반으로 구현하였다. 전체 시스템은 크게 사용자 프로파일, 사례데이터베이스, 유사 사례 검색, 개인별 콘텐츠 추천, 사용자 프로파일 수정 부분으로 구성된다.

제안한 추천시스템 모델에 대한 효율성 검증을 위하여 Recall, Precision, F-measure 등과 같은 측정 기법에 의해 협동적필터링 기법의 추천 효율과 비교 실험하였다 [4, 5, 6]. 실험결과 유사 사용자의 평가정보를 이용한 협동적필터링 기법보다 효율적인 개인화 전략이 가능함을 보였다.

### 2. 관련연구

#### 2.1. 사례기반추론

사례기반추론기법은 인간의 문제해결 과정처럼 과거에 이미 사용했던 문제 해결 방식을 사용하여 현재의 문제 해결에 이용하는 방법이다. 이 기법은 새로운 문제가 발생하면 검색과정에서 가장 유사한 사례를 찾아내고 적응 과정에서 이전 사례의 해를 재 사용하여 문제를 해결한

본 연구는 정보통신부 대학기초연구지원사업(과제번호 : 2001-111-2)에 의해 수행되었음.

다. 또한 교정과정에서 새로운 문제의 해가 문제 해결에 적합한지를 검증, 교정한 후 새로운 사례로 저장한다. 사례는 과거의 경험에 당시 상황 및 문제 해결 방법을 표현한 것이다. 사례의 속성은 사례의 검색을 위해 사용하는 인덱스 속성과 문맥상의 정보를 제공하는 비 인덱스 속성이 있다. 사례의 검색(Retrieval)은 인덱싱 방법과 밀접한 관련이 있으며 과거 사례에서 유사성이 높은 사례를 찾는 과정이다. 적용은 새로운 문제에 대해 검색된 사례와의 차이를 고려하여 새로운 해를 적용시키는 과정이다[7]. 다음 그림 1은 사례기반추론 기법의 개념도이다.

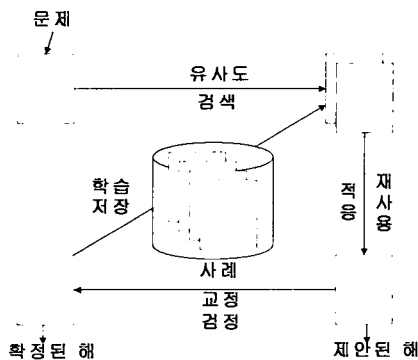


그림 1. 사례기반 추론 시스템 개념도

### 3. 제안 추천 시스템

#### 3.1. 제안 시스템의 구성

제안한 추천시스템은 사용자 프로파일, 사례데이터베이스, 유사 사례 검색, 채용정보 추천, 사용자 프로파일 수정 부분으로 구성된다. 다음 그림 2는 제안 시스템의 전체 구성도이다.

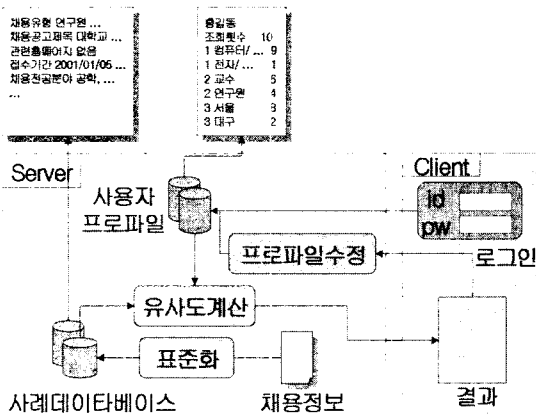


그림 2. 제안시스템 구성도

#### 3.2. 채용정보의 사례 데이터베이스 저장

실험에서 사용된 채용정보는 사례의 형태로 표준화된다. 사례에서 채용전공분야, 채용유형, 희망근무지는 사용자들이 콘텐츠를 결정하는 중요 속성이므로 인덱스 속성이고 나머지 접수기간, 관련 홈페이지, 급여조건 등은 비 인덱스 속성이다.

#### 3.3. 사용자 프로파일 구성

사용자 프로파일은 콘텐츠에 대한 개인별 선호도를 나타내고 목시적인 평가정보에 의해 수집된다. 사용자 프로파일을 구성함으로써 사용자의 속성과 속성 값에 대한 관심의 정도를 알 수 있다. 제안 시스템에서 사용자의 초기 이용 시 목시적인 평가정보의 미흡으로 인해 사용자 프로파일 정보가 부족할 때 추천은 사용자가 회원가입 할 때 등록한 기본정보를 이용하여 이루어진다.

#### 3.4. 유사도 기반의 사례검색

속성 유사도( $f(T_i, S_j)$ )는 입력문제와 사례 사이에 일대일 대응하는 속성에 포함 된 속성 값간의 유사도이다. 속성 값 가중치( $W_{sj}$ )는 각 사례에 포함되어 있는 속성의 속성 값에 대한 사용자의 관심정도를 의미한다. 다음 식 1은 본 제안 시스템에서 유사도를 구하는 공식이다.

$$Similarity(T, S) = \sum_{i=1}^n f(T_i, S_j) \times W_{sj} \times W_i \quad (1)$$

여기서, T는 입력문제, S는 사례, n은 각 사례의 속성의 수, i는 1부터 n 까지 각각의 속성을 의미한다.

속성 가중치( $W_i$ )는 기준 값과 각 속성 값의 거리 합이며, 다음 식 2로 표현한다. 기준 값은  $1/(속성\ 값\ 개수)$ 로 정의한다.

$$W_i = \sqrt{\sum_{j=1}^n (a_{ij} - b_j)^2} \quad (2)$$

여기서,  $a_{ij}$ 는 속성  $a_i$ 의 j번째 속성 값의 속성 값 윗수를 조회횟수로 나눈 값을 의미한다.  $b_j$ 는 속성  $a_i$ 의 기준 값이다.

#### 3.5. 사용자 프로파일 수정

사용자 프로파일은 추천된 채용정보 중에서 사용자가 채용 정보를 열람할 때 남겨지는 로그 정보를 분석하여 개인별로 수집한다. 수집된 로그 정보는 추천시 사용된다.

### 4. 실험 결과 및 고찰

#### 4.1. 실험 환경

제안한 추천 기법의 효율성을 검증하기 위하여 채용정보사이트의 사용자 기록을 사용하였다. 시스템은 UNIX에서 오라클 DBMS 8.1.7을 기반으로 구현하였으며 사용언어는 Servlet과 JSP이다.

실험 데이터는 사용자 65명에게 307개의 콘텐츠를 열람하게 한 후 콘텐츠에 대한 선호도를 1에서 5의 크기로 명시적으로 평가하게 하였고 동시에 사용자의 접속 로그를 분석하여 묵시적 평가정보로 사용하였다. 실험 데이터는 표2 와 같다.

표 2. 실험 데이터

항목	사용자	콘텐츠	묵시적	명시적
개수	30	250	957	957

4.2. 성능 평가 기준

Recall은 사용자가 선호하는 콘텐츠 중에서 얼마나 많은 콘텐츠가 추천이 되었는지 나타내는 평가 방법으로 식 (3)로 정의한다.

$$Recall = \frac{\text{정확하게 추천된 콘텐츠 수}}{\text{사용자가 선호하는 콘텐츠 수}} \quad (3)$$

Precision은 추천 리스트 중에서 몇 개의 콘텐츠를 사용자가 실제로 좋아했는지를 나타내는 평가방법으로 식 (4)로 정의한다.

$$Precision = \frac{\text{정확하게 추천된 콘텐츠 수}}{\text{추천 리스트의 콘텐츠 수}} \quad (4)$$

F-measure는  $\beta$ 값에 따라 Recall과 Precision의 중요도를 부여하여 평가하는 방법이다[6].

$$F_{\beta} = \frac{(\beta^2 + 1) \times precision \times recall}{\beta^2 \times precision + recall} \quad (5)$$

4.3. 실험 결과 및 분석

다음 표3은 본 논문에서 제안기법과 협동적필터링 기법의 비교 실험 결과를 나타낸다. 횡수마다 랜덤하게 실험 값을 선택하여 실제 명시적 평가 값과 비교하였다.

표 3. 추천 리스트에 대한 비교 실험 결과

횡수	제안기법			협동적필터링		
	Recall	Precision	F-Measure	Recall	Precision	F-Measure
1	0.71	0.98	0.82	0.71	0.83	0.76
2	0.86	0.75	0.8	0.57	0.80	0.67
3	0.98	0.78	0.87	0.71	0.71	0.71
4	0.67	0.67	0.67	0.33	0.98	0.49

실험 결과, 제안기법이 협동적 필터링 기법의 Recall, Precision, F-Measure 값 보다 모두 높은 값이 나왔음을 알 수 있다.

5. 결론

본 논문에서는 인터넷 사이트에서 개인별로 콘텐츠를 효과적으로 추천하기 위한 추천시스템 모델을 제안하였다. 제안한 모델은 묵시적인 평가정보를 이용한 사례기반추론 기법에 의해 구현하였다.

제안한 기법에 대한 Recall, Precision, F-measure의 평가 방법을 통해 실험결과 협동적필터링 기법 보다 모든 부분에서 더 좋은 결과가 나왔음을 볼 수 있다. 그러므로 제안 시스템이 유사 사용자의 평가정보를 이용한 협동적필터링 기법보다 효율적인 개인화 전략이 가능하다고 할 수 있다.

제안한 추천기법은 유사 집단의 추천을 필요로 하지 않을 뿐만 아니라 상품 구매속성을 추천에 반영할 수 있다. 그리고 사용자가 많아져도 복잡도가 증가하지 않으며 새롭게 등록된 콘텐츠도 추천할 수 있다. 또한 제안한 추천기법은 사용자 프로파일에 의해 개인화 된 추천이 가능하며 묵시적 가중치를 이용하므로 사용자가 속성에 대한 가중치를 명시적으로 지정해야하는 불편함을 보완하였다. 앞으로 본 제안 모델을 이용하여 일대일 마케팅을 위한 eCRM시스템 개발이 가능하리라 예상된다.

참고문헌

- [1] Keith Bradley, Rachael Rafter, Barry Smyth, "Personalised Case Retrieval," In Proceedings of 10th Irish Conference on Artificial Intelligence and Cognitive Science, Cork, Ireland, September, 1999
- [2] 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, April, 2000
- [3] Douglas W. Oard, Jinmook Kim, "Implicit Feedback for Recommender Systems," In Proceedings of AAAI Workshop on Recommender Systems, pp.80-82, Madison, Wisconsin, USA, July 1998
- [4] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom and John Riedl, "GroupLens: An Open Architecture for Collaborative Filtering of Netnews," In Proceedings of the ACM 1994 Conference on Computer Supported Cooperative Work, pp.175-186, 1994
- [5] Michael Buckland, "The Trade-off between Recall and Precision," Journal of the American Society for Information Science, pp.12-19, 1994
- [6] Lewis, D. and Gale, W.A., "A Sequential Algorithm for Training Text Classifiers," In Proceedings of the 17th Annual International ACM-SIGIR Conference on Research and Development in Information Retrieval, London, Springer-Verlag, pp.3-12, 1994
- [7] Ian Watson, "Applying Case-based Reasoning: Techniques for Enterprise Systems," Morgan Kaufmann Publishers, Inc, Chapter 2, 3, 1997