

## 콘텐츠 추천 시스템의 객관적 성능평가 지원을 위한 정확도 평가 시스템 설계 및 구현

\*김다희 \*\*신사임 \*\*\*박성주

전자부품연구원

\*daheiy@naver.com

## Design and Implementation of the Evaluation System for the User-based Contents Recommendation Systems

\*Kim, Da-Hee \*\*Shin, Sa-Im \*\*\*Park, Sung-Joo

Korea Electronics Technology Institute, Digital Media Research Center

## 요약

추천 엔진의 지속적인 성능 향상을 위해서는 객관적 성능 평가가 이루어져야 하지만, 콘텐츠 추천 기술은 평가 데이터 구축이 어렵고 평가를 위한 심도 깊은 연구가 이루어지지 못하여 많은 어려움을 가지고 있다. 본 논문은 콘텐츠 추천 시스템의 객관적인 평가를 시스템 구축에 대한 연구이다. 추천 알고리즘의 효율적이고 객관적인 성능 평가를 위하여 기존 연구들을 분석하여 대표적인 성능평가 양식들을 구현하였다. 통계적인 평가를 위한 사용자 데이터를 수집하였으며 데이터 크기, 평가방식, 추천 알고리즘의 모듈 별 성능 추이를 쉽게 변경하고 관찰할 수 있도록 인터페이스를 설계하였다. 이러한 평가 시스템의 도입으로 콘텐츠 추천 알고리즘의 지속적인 성능 보안을 기대 할 수 있을 것이다.

## 1. 서론

Web 2.0시대로 들어서면서 웹 관련 콘텐츠 서비스들이 많이 상용화되었다. 사용자가 콘텐츠를 자유롭게 생성, 가공, 공급하는 서비스가 되면서 웹에서 많은 콘텐츠들을 쉽게 접하게 되었다 [1]. 하지만 많은 콘텐츠가 웹에 존재하다보니 사용자는 자신이 원하는 콘텐츠를 웹에서 찾는 것이 어려워졌다. 그래서 사용자들에게 필요한 콘텐츠를 추천해주는 추천 시스템들이 많이 개발되고 있는 추세이다.

콘텐츠 추천 시스템이란 사용자 선호도 분석 및 예측 기술을 기반으로 사용자에게 최적의 콘텐츠를 추천하는 기술이다 [6].

이러한 추천 시스템들의 성능 평가를 위해서는 평가를 위한 데이터 구축이 되어야하고 사용자들의 개인 정보들을 수집하는 작업이 필요하다. 하지만 이러한 데이터 구축 및 개인 정보 수집이 어렵기 때문에 추천 시스템을 객관적으로 평가하기가 어려웠다. 또한, 공통적으로 광범위하게 사용되는 평가방식은 아직 없다.

본 논문은 기존의 연구 평가를 위한 방법들을 한 틀에 모아 평가하고 알고리즘의 변화에 따라서 객관적인 평가가 바로 가능한 시스템을 구현하고자 했다. 따라서 객관적인 평가가 가능한 평가 시스템이 구현된다면 추천 시스템의 성능은 향상 될 것이다.

## 2. 콘텐츠 추천 엔진의 평가 방식

콘텐츠 추천에 관한 기존 연구들을 분석해본 결과 이 분야에서 객관성을 보장할 만한 대표적인 평가 방법은 확립되지 않은 상태이다. 그러므로 본 연구에서는 많은 기존 연구의 분석을 통하여 콘텐츠 추천 시스템에서 사용되었던 기존 평가 방식들을 수집하여 보았고, 다음과

같다.

## 가. 정확률(Precision) 및 재현율(Recall)

사용자  $a$ 가 실제 이용한 콘텐츠의 집합을  $V_a$ , 사용자에게 추천된 콘텐츠의 집합을  $R_a$ , 추천된 콘텐츠 가운데 사용자가 실제로 이용한 콘텐츠의 집합을  $|V_a \cap R_a|$  라고 하고  $N_u$ 은 추천받는 사용자의 전체 수라고 하면, 정확률과 재현율은 각각 (2-1), (2-2)와 같이 표현된다 [2].

$$Precision = \sum_{i=1}^{N_u} \frac{|V_a \cap R_a|}{|R_a|} \quad (2-1)$$

$$Recall = \sum_{i=1}^{N_u} \frac{|V_a \cap R_a|}{|V_a|} \quad (2-2)$$

## 나. 평균 절대 오류(MAE, Mean Absolute Error)

추천에 대한 오류는 실제 선호도 값과 예측된 선호도 값과의 차이로 정의되며,  $MAE$ 는 이러한 오류의 절대 값의 평균을 의미한다.  $MAE$ 는 절대적으로 추천 시스템이 얼마나 정확하게 추천 했는지를 나타낸다.  $R_{a_i}$ 와  $V_{a_i}$ 는 각각 추천된 콘텐츠와 사용자가 선택한 콘

텐츠의 순위이고,  $N$ 은 추천된 콘텐츠의 전체 개수이다 [2].

$$MAE = \frac{\sum_{i=1}^N |R_{a_i} - V_{a_i}|}{N} \quad (2-3)$$

### 다. 커버리지 (Coverage)

전체 콘텐츠 가운데 추천 시스템이 추천할 수 있는 콘텐츠의 비율이다. 즉, 커버리지는 목표 사용자 또는 콘텐츠 집합을 전체로 보았을 때, 추천 시스템이 예측할 수 있는 사용자 혹은 콘텐츠 개수의 전체 대비 비율을 의미한다 [3]. 수식으로 표현하면 (2-4)와 같다.

$$Coverage = \frac{R_a}{N} \quad (2-4)$$

### 라. 민감도 (Sensitivity)와 특이성 (Specificity)

민감도는 추천 시스템에 의해서 추천된 콘텐츠 가운데 사용자가 실제로 선호한 콘텐츠의 비율을 의미하고, 특이성은 추천되지 않은 콘텐츠 가운데 사용자가 실제로 선호하지 않은 콘텐츠의 비율을 의미한다. 사용자가 선호한 콘텐츠는  $P$ 이고, 선호하지 않은 콘텐츠는  $N - P$ 이다. 수식으로 표현하면 (2-5), (2-6)과 같이 표현 할 수 있다.

$$Sensitivity = \frac{P}{R_a} \quad (2-5)$$

$$Specificity = \frac{N - P}{N - R_a} \quad (2-6)$$

### 마. 역순위 평균 (MRR, Mean Reciprocal Rank)

역순위 평균은 검색 시스템에 요청한 질의의 실제 정답을 포함하는 문서가 시스템이 제시한 검색 결과에서 몇 위에 위치하는지를 파악하는 것이다.  $i$ 는 데이터 셋에 각각의 콘텐츠를 의미하고,  $R_{a_i}$ 는 추천 방송 목록에서  $i$ 번째 순위를 의미한다 [4].

$$MRR = \frac{1}{N} \sum_{i=1}^N \frac{1}{R_{a_i}} \quad (2-4)$$

### 바. Top-N-Hit-Rate

실제 추천 콘텐츠의 리스트를 1위부터 내림차순으로 사용자에게 제공할 때, 사용자가 원하는 콘텐츠가 N위에 있는 비율을 의미한다 [7].

콘텐츠 추천 알고리즘을 사용자 정보 취득 방식에 따라 분류하면 명시적 사용자 정보 취득 방식과 암묵적 사용자 정보 취득 방식이 있

다. 명시적 사용자 정보 취득 방식은 콘텐츠에 대해서 사용자가 명시적으로 등급 정보를 부여하게 되고, 이 정보를 기반으로 등급이 없는 콘텐츠들에 대한 추천을 수행한다. 암묵적 사용자 정보 취득 방식은 추천 엔진이 간접적으로 사용자의 콘텐츠와 관련된 행동, 소비 등의 히스토리를 획득하여 사용자가 선호할 것으로 예상되는 콘텐츠를 추천한다.

명시적 사용자 정보 기반 추천 시스템은 사용자들이 콘텐츠의 점수를 표기하여 추천 시스템은 사용자가 표기한 점수를 갖고 추천 콘텐츠를 선별하였고, 기존의 추천 시스템은 MAE나 MRR로 추천 시스템을 평가를 할 수 있었다. 본 논문에서 평가하는 추천 시스템은 암묵적 사용자 정보 기반의 추천 시스템이기 때문에 기존의 평가 방식에서 MAE나 MRR을 제외한 정확률 및 재현율, 커버리지, 민감도 및 특이성, Top-N-Hit-Rate를 중점으로 평가 방식을 구현하였다.

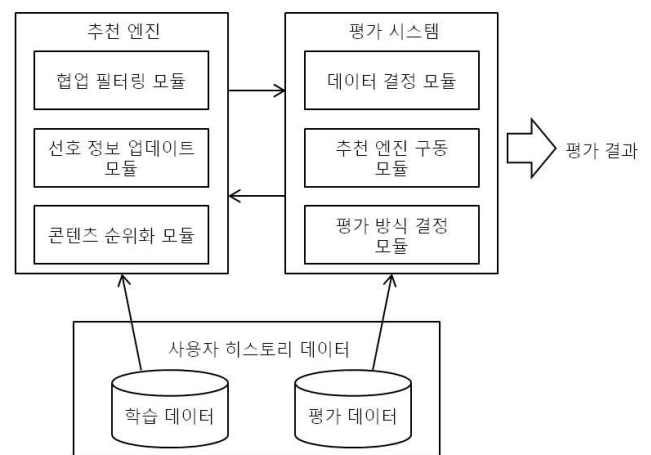


그림 1 평가 시스템 구현을 위한 시스템 흐름도

## 3. 평가 시스템 구현

사용자 히스토리 데이터를 이용하여 추천 엔진과 평가 시스템을 구현한다. 평가 시스템 사용자가 원하는 결과를 위하여 평가 시스템에서 데이터 결정 모듈과 추천 엔진 구동 모듈을 선택하여 추천 엔진에서 추천 데이터를 얻을 수 있다. 얻어진 추천 데이터와 사용자 히스토리 데이터 중 평가 데이터를 사용하여 평가 방식 결정 모듈을 통해 평가 결과를 얻을 수 있다.

### 가. 사용자 히스토리 데이터

추천 시스템을 평가에 사용되는 데이터를 구축하기 위해 사용자가 사용한 데이터들을 저장 할 수 있는 사용자 히스토리 수집 틀을 구현하였다. 사용자 히스토리 수집 틀에 저장된 사용자 히스토리를 이용하여 평가에 필요한 데이터를 얻었다. 사용자 히스토리 수집 틀에서 가장 콘텐츠 양이 많은 음악과 방송 콘텐츠를 평가 데이터로 사용하였다. 표 3-1를 통하여 추천 시스템에 쓰인 데이터들을 확인 할 수 있다.

구분	방송 콘텐츠	음악 콘텐츠
사용자 수	47	49
콘텐츠 개수	491	2,122
사용자들이 이용한 전체 콘텐츠 개수(누적)	3,780	8,494
사용자별 평균 콘텐츠 이용 개수	80.43	173.35

표 3-1 추천 시스템 평가를 위한 콘텐츠 정보

사용자의 사용 히스토리의 개수가 방송과 음악 각각 50개를 기준으로 50개 이상일 때와 미만일 때로 나누어 평가를 하였다. 50개 이상 데이터 셋은 추천엔진이 평가 가능한 충분한 데이터 수를 확보하고 있는 셋이다. 전체 데이터를 포함하고 데이터 셋은 모든 사용자들의 평가가 가능하고 협업 필터링의 엔진의 성능을 평가하기에 적합한 데이터 셋이다.

사용자의 사용 히스토리를 7:3의 비율로 나누어 학습 데이터와 평가 데이터로 정의하였다. 추천 시스템에서 학습 데이터를 이용하여 추천한 결과 데이터와 평가 데이터를 이용하여 평가를 하였다.

### 나. 평가 시스템 인터페이스

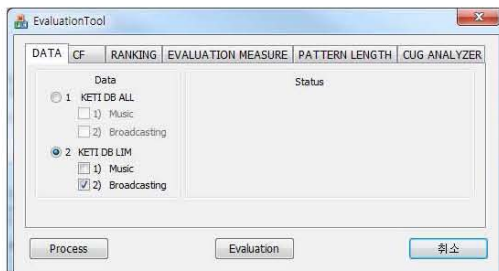


그림 1 데이터 결정 모듈 인터페이스

사용자 히스토리의 개수를 기준으로 모든 사용자를 포함하는 ALL 데이터와 일부 사용자만을 포함하는 LIM 데이터로 나누었다. 음악 콘텐츠와 방송 콘텐츠를 선택하여 성능 평가를 할 수 있도록 하였다.

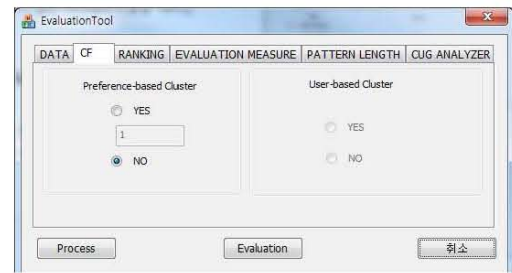


그림 2 추천 엔진 구동 모듈 인터페이스 - 협업기반 선택

협업 기반 선택 시트에서는 평가에 필요한 평가 방식을 선택하여 결과 도출한다 [7].

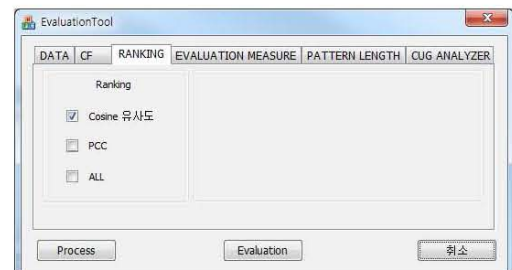


그림 3 추천 엔진 구동 모듈 인터페이스 - 순위화 방법 선택

순위화 방법 선택 시트에서는 추천 시스템에서 순위 정보를 도출하기 위해 쓰이는 코사인 유사도와 PCC( Pearson's Correlation Coefficient)를 선택할 수 있게 되어있다 [5,8].

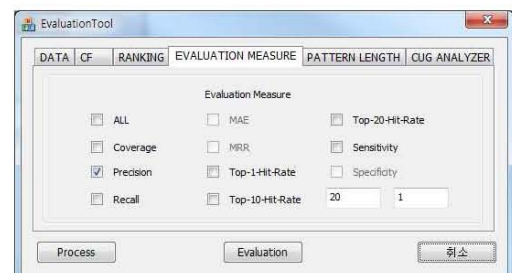


그림 4 평가 방식 결정 모듈

EVALUATION MEASURE 시트에서는 평가에 필요한 평가 방식을 선택하여 결과 도출한다.

### 다. 평가 시스템 구동 결과

아래 그림에서는 평가 시스템을 구동한 결과의 예를 보여주고 있다. 평가 시스템의 각 시트에서 원하는 데이터를 선택하고 Process 버튼을 누르면 추천 시스템에서 추천한 추천 데이터 파일이 만들어진다. 그 뒤 Evaluation Measure 시트에서 원하는 평가 방법들을 선택하고 Evaluation 버튼을 누르면 원하는 결과들이 시트 아래 표시된다.

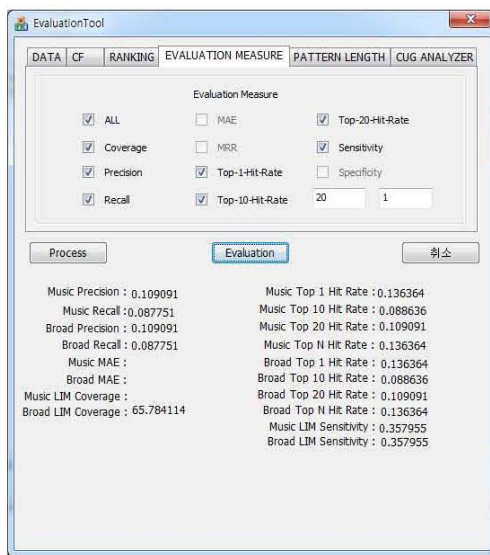


그림 6 평가 시스템 구동의 예

#### 4. 결론

본 논문은 명시적 사용자 기반 추천 시스템들과 다른 암묵적 사용자 기반 추천 시스템을 평가하기 위한 평가 시스템 구현 연구이다. 평가 시스템에서 사용된 방식은 광범위하게 사용되는 평가 방식들이 아니라 기존 연구 분석을 통해 암묵적 사용자 기반 추천 시스템에 맞게 평가 방식을 구현하였다. 보다 손쉽게 평가 할 수 있도록 지원하였고 추천 시스템의 성능 향상에 기여할 수 있을 것이다.

본 연구는 지식경제부 산업융합원천기술개발사업의 지원을 받아 수행된 연구결과임. (No. 10031824)

#### 5. 참고문헌

- [1] 김재하 “웹 2.0 환경의 새로운 디지털 콘텐츠 서비스 전망에 관한 연구”, 한국방송공학회
- [2] Sa-im Shin, Jong-Soel Lee, Soek-Pil Lee, Tae-Boem Lim “Automatic management of the user preferences with the information retrieval techniques”, ICUT 2007
- [3] H. Kwon and K. Hong, “Personalized smart TV program recommender based on collaborative filtering and a novel similarity method,” Proceedings of IEEE Transaction on Consumer Electronics, Vol. 57, No. 3, pp. 1416-1423, 2011.
- [4] E. Kim, “An efficient collaborative content recommendation technique for smart TV environments,” Ph. D. thesis, Department of Computing, Graduate School of Soongsil University, 2011.
- [5] 신사임, 이종철, 이석필 “사용자 패턴의 자동 추출을 통한 TV-Anytime 기반 사용자 선호 정보 관리 시스템” 2006 정보과학회
- [6] 박성주 “이용 패턴 정보 기반 멀티미디어 콘텐츠 추천 시스템 연구”, 광운대학교
- [7] MIREX “[http://www.music-ir.org/mirex/wiki/MIREX\\_HOME](http://www.music-ir.org/mirex/wiki/MIREX_HOME)”
- [8] 김진상, 윤병주 “협동적 필터링을 위한 동시출현빈도 사용의 제한 피어슨 알고리즘”, 한국정보처리학회