# Implicit Feedback을 통한 선호도 예측 알고리즘 구현

*장정록, 김용구, 김도연
삼성전자 디지털 미디어 연구소

# Implementation Of User Preference Estimation Algorithm Using Implicit Feedback

*Jeongrok Jang, Yongu Kim, Doyeon Kim
Digital Media R&D Center
Samsung Electronics

## Abstract

In this paper, we propose a new approach for the implicit rating algorithm of finding user's intense and preference to the contents on the web. Although the explicit method dig out the user preference of specific contents based on the user's intervention, we propose the implicit method obtaining the user preference according to the user's behavioral patterns on the web implicitly and automatically without the user's intervention. The implementation results show that the proposed approach is highly valuable for supporting recommender systems in conjunction with the users lifestyle.

## I. 서론

일반적으로 사용자의 선호도를 파악하는 방법에는 명시적인 방법(Explicit Method)과 묵시적인 방법(Implicit Method)이 있다[1]. 명시적인 방법은 사용자가 직접 평가를 바탕으로 선호도를 파악하는 방법으로, 사용자의 입력이 없는 데이터에 대해서는 선호도를 파악할 수 없다는 문제점이 있다. 묵시적인 방법은 사용자의 행동양식을 이용하여 선호도를 추출하는 방법으로, 특히 사용자의 선호도를 사용자의 입력 없이 자동으로 평가할 수 있어 명시적인 방법의 단점을 보완할 수 있다[2].

본 논문에서는 웹 컨텐츠에 대한 사용자의 선호도를 평가하는 방법에 있어서, 사용자의 행동양식과 명시적인 평가정보와의 연관성을 찾는 부분에 초점을 맞추어 사용자의 선호도 예측 알고리즘 구현 기법을 제안하고 효율을 측정하였다.

## II. 본론

### 2.1 데이터 수집과 분석

묵시적인 평가 방법은 사용자의 묵시적인 행동패턴을 수집, 분석한 데이터를 사용자의 선호도를 평가하는 정보로 사용하는 방법을 말한다[3]. Internet Explorer를 사용하는 환경에서 수집할 수 있는 사용자의 묵시적인 행동패턴은 마우스와 키보드 등 인풋 장치를 사용하는 과정에서 발생하는 것과 인쇄, 저장 등의 행동, 그리고 해당 Web 컨텐츠에서 머문 시간 등 컨텐츠에서 추출할 수 있는 정보가 있다.

### 2.2 데이터 분석

사용자의 행동패턴과 명시적인 평가정보 사이의 연관성을 파악하기 위해서, 묵시적인 방법으로 수집한 데이터를 분석해서 사용자가 컨텐츠에 명시적으로 평가한 선호도와 연관성을 파악했다. 수집된 데이터를 통해 연관성을 파악하기 위해 Minitab, Matlab, Excel 등의 Tool을 이용하여 사용자 행동패턴을 분석하였다.

## Ⅲ. 구현

Internet Explorer 환경에서 사용자의 명시적인 평가

정보와 묵시적인 데이터 수집을 위해서 Browser에 애드인 되는 툴바 프로그램을 그림1과 같이 구현했고, 해당 프로그램의 구조는 그림2에 나타나 있다. 툴바 프로그램은 버튼을 이용해서 사용자로부터 명시적인 평가정보를 5단계로 나누어서 수집한다. 또한 웹 컨텐츠를 소비하면서 나타내는 사용자의 행동패턴을 메시지 후킹 방식을 이용해서 수집한다.

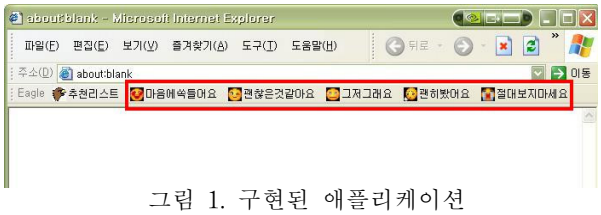


그림 1. 구현된 애플리케이션

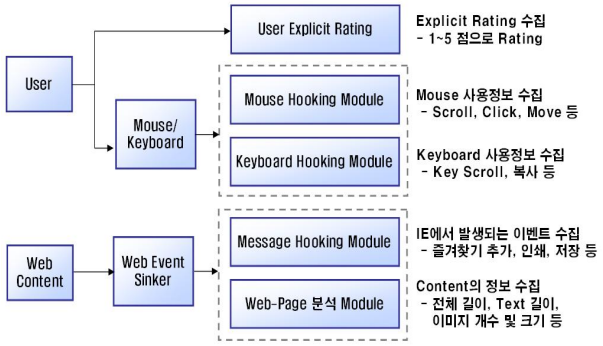


그림 2. 전체 구조도

수집하는 데이터의 종류는 마우스를 이용한 클릭횟수, 스크롤 횟수, 움직임 횟수와 키보드 방향키를 이용한 스크롤 횟수가 있다. 또한 사용자의 웹 컨텐츠 소비시간, 전체 컨텐츠길이에서 사용자가 본 컨텐츠 길이의 비율, 컨텐츠의 텍스트 길이 등 컨텐츠와 관련된 데이터와 즐겨찾기 추가, 인쇄, 저장 등의 사용자의 행동 패턴과 관련된 데이터를 수집한다.

50명의 사용자에게 해당 프로그램을 배포하고, 1달 정도 사용하도록 하여 총 4000여개의 데이터를 수집하였다. 이중 Web Text가 아니거나 해당 Web Text에서 아무런 동작도 취하지 않은 것과 같은 무효 데이터를 필터링하여 최종적으로 2972개의 유효 데이터를 수집했다. 표1은 수집한 데이터의 사용자의 선호도 분포를 나타내고 있다.

해당 데이터를 바탕으로 Matlab, Minitab, Excel 등의 툴을 사용해서 명시적인 평가정보와 사용자의 묵시적인 행동패턴과의 상관관계에 대해서 분석했다. 사용자의 행동패턴 중, 컨텐츠 소비 시간, 컨텐츠 길이 당 사용자가 머문 시간, 마우스 무브 등 3개의 인자가 가장 연관성이 높은 것으로 분석이 되었고, 그에 따른 회귀식이 도출되었다.

| | |
|---|---|
| Explicit Rating Level = 5 | 514 |
| Explicit Rating Level = 4 | 1280 |
| Explicit Rating Level = 3 | 906 |
| Explicit Rating Level = 2 | 225 |
| Explicit Rating Level = 1 | 47 |
| TOTAL | 2972 |

표 1. 사용자의 Explicit Rating Level

사용자가 직접 컨텐츠에 대한 선호도를 명시적으로 평가한 데이터와 회귀식에서 계산된 결과 값을 RMSE (Root Mean Square Error)식을 이용해서 비교했다. RMSE의 의미는 통계학에서 표준편차의 의미와 비슷한 개념이다. 즉, 예상한 값과 실제 관측한 결과 값의 평균적인 차이이다. 계산된 RMSE의 값은 0.8198의 값이 나왔다. 세계적 수준의 영화 추천 시스템인 Cinematch의 RMSE 값 0.9514와 간접 비교를 통해 본 논문에서 구현한 알고리즘의 성능이 탁월하다고 할 수 있다. 향후, 추천 시스템에서 사용자의 성향을 파악하는 분야에서, 본 알고리즘이 활용될 수 있을 것으로 본다.

## Ⅳ. 결론 및 향후 연구 방향

본 논문에서는 Internet Explorer 환경에서, 사용자의 묵시적인 행동패턴을 바탕으로 컨텐츠에 대한 선호도를 파악하는 알고리즘 구현에 목적을 두었다. 사용자가 Internet Explorer를 사용하면서 나타내는 행동양식을 수집 분석해서, 사용자가 명시적으로 평가한 선호도와의 연관성을 추출하는 과정을 통해 선호도 예측 알고리즘을 구현했다. 본 논문에서 구현한 사용자의 묵시적인 행동패턴을 바탕으로 선호도를 예측하는 알고리즘을 응용해서, 향후 추천 시스템에서 사용자의 프로파일 생성에 활용될 수 있을 것으로 예상한다.

## 참고문헌

[1] Mark Claypool, David Brown, Phong Le, and Makoto Waseda, "Inferring User Interest", Internet Computing, IEEE, Vol 5, Issue 6, Nov-Dec 2001, pp. 32-39.

[2] D.M Nichols, "Implicit Rating and Filtering", Proc. 5th DELOS Workshop on Filtering and Collaborative Filtering, Nov. 1997, ERCIM, Sophia Antipolis, France, pp. 31-36

[3] J. Kim,D.Oard, and K. Romanik, "User Modeling for Information Filtering Basedon Implicit Feedback," Proc. ISKO-France, ISKO, Paris, 2001