



## 협력 필터링 기반 게임 추천 시스템

Game Recommendation System Based on Collaborative Filtering

---

저자 (Authors)	황영웅, 김종현, 김병만, 이현아 Young-Woong Hwang, Jong-Hyun Kim, Byeong-Man Kim, HyunAh Lee
출처 (Source)	<a href="#">한국정보과학회 학술발표논문집</a> , 2017.06, 1914-1916 (3 pages)
발행처 (Publisher)	<a href="#">한국정보과학회</a> KOREA INFORMATION SCIENCE SOCIETY
URL	<a href="http://www.dbpia.co.kr/Article/NODE07207799">http://www.dbpia.co.kr/Article/NODE07207799</a>
APA Style	황영웅, 김종현, 김병만, 이현아 (2017). 협력 필터링 기반 게임 추천 시스템. 한국정보과학회 학술발표논문집, 1914-1916.
이용정보 (Accessed)	고려대학교 163.***.133.25 2017/09/06 16:16 (KST)

---

### 저작권 안내

DBpia에서 제공되는 모든 저작물의 저작권은 원저작자에게 있으며, 누리미디어는 각 저작물의 내용을 보증하거나 책임을 지지 않습니다. 그리고 DBpia에서 제공되는 저작물은 DBpia와 구독 계약을 체결한 기관소속 이용자 혹은 해당 저작물의 개별 구매자가 비영리적으로만 이용할 수 있습니다. 그러므로 이에 위반하여 DBpia에서 제공되는 저작물을 복제, 전송 등의 방법으로 무단 이용하는 경우 관련 법령에 따라 민, 형사상의 책임을 질 수 있습니다.

### Copyright Information

Copyright of all literary works provided by DBpia belongs to the copyright holder(s) and Nurimedia does not guarantee contents of the literary work or assume responsibility for the same. In addition, the literary works provided by DBpia may only be used by the users affiliated to the institutions which executed a subscription agreement with DBpia or the individual purchasers of the literary work(s) for non-commercial purposes. Therefore, any person who illegally uses the literary works provided by DBpia by means of reproduction or transmission shall assume civil and criminal responsibility according to applicable laws and regulations.

# 협력 필터링 기반 게임 추천 시스템

황영웅<sup>○</sup> 김종현 김병만 이현아  
금오공과대학교 컴퓨터소프트웨어공학과

woong9423@gmail.com, fussaep@gmail.com, bmkim@kumoh.ac.kr, halee@kumoh.ac.kr

## Game Recommendation System Based on Collaborative Filtering

Young-Woong Hwang<sup>○</sup> Jong-Hyun Kim, Byeong-Man Kim, HyunAh Lee  
Dept. of Computer Software Engineering, Kumoh National Institute of Technology

추천 시스템은 급격히 증가하고 있는 인터넷 정보 중 원하는 정보만을 추출하는데 유용하다. 추천 시스템에는 내용 기반 추천 시스템과 협력 필터링 기반 추천 시스템, 이 둘의 장점을 혼합한 추천 시스템 등이 존재한다. 이 중 협력 필터링은 텍스트 기반의 정보가 부족한 분야에서의 추천 시스템을 구성하는데 많이 쓰인다. 게임은 텍스트 기반의 정보를 적용하기 어려운 장르 중 하나로, 협력 필터링을 적용하기에 적합하다. 본 논문에서는 협력 필터링을 이용한 게임 추천 시스템을 개발한다. 구축된 시스템에서는 보완 유사도 측정 방법과 평가항목 중복도를 기준으로 이웃을 선택하여, 4.1초의 처리시간에서 평균 정확도 61.2%와 최고 정확도 71.7%를 얻어 빠른 시간 내에 우수한 추천 결과를 얻을 수 있었다.

### 1. 서 론

게임 산업의 발전으로 인하여 게임 종류의 다양화와 함께 게임 이용자의 연령이나 특성도 다양해지고 있다. 이에 맞춰 게임사들은 새로운 게임을 선보이고 있어 게임의 종류와 수는 크게 증가하고 있다. 게임 이용자들은 자신이 구매하려는 게임이 적절한 가격 대비 성능을 가지는지 파악하기 위해서 공신력 있는 게임 전문 웹진이나, 게임 전문 스트리머의 방송, 주변 지인의 추천에 주로 의존하는데, 이러한 정보들에는 광고성 기사나 방송이 상당수 포함되어 정확성이 떨어진다. 또한 자신의 취향을 정확히 반영한 게임 추천을 받기 어렵다는 단점이 존재하여, 데이터 기반의 적합한 게임 추천 시스템이 요구된다.

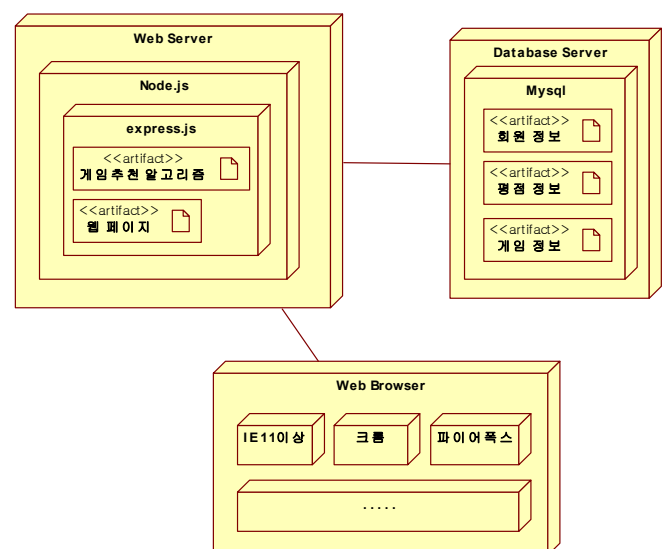
추천 시스템을 구성하는데 많이 활용되는 방식으로는 협력 필터링 방식과 내용기반 필터링 방식이 있다. 내용기반 필터링을 게임 추천에 적용하기 위해서는 사용자 프로파일링과 같은 적합한 대용량의 텍스트 정보를 얻기 어려운 문제가 있다. 이에 반하여 협력 필터링 방식은 적합한 데이터를 획득하기는 쉬우나 데이터의 양이 늘어날수록 계산 속도가 떨어지는 문제를 가지고 있어, 웹서비스에서 중요시 되는 확장성에서 치명적인 단점을 가진다. 또한, 초기데이터의 부족으로 인한 정확성 저하의 문제, 즉 콜드 스타트 문제가 존재한다.

김병만 등[1]은 협력 필터링의 문제점 중 콜드 스타트 문제를 해결하기 위해 사용자의 프로파일을 클러스터링하여 내용기반 필터링을 적용하는 방법과 사용자간 평가 데이터에서 계산된 유사도를 이용하는 협력 필터링 두 가지를 혼합하여 사용하였다. 본 연구에서는 [1]의 접근을 게임 분야에 적용하되, 획득하기 어려운 사용자 프로파일을 제외하고 사용자간 평가 데이터만을 사용하여 협력 필터링 기반 추천 시스템을 구축한다. 협력 필터링의 문제점인 콜드 스타트 문제를 해결하기 위해서는 게임

전문 사이트에 축적된 게임 평가 데이터와 각 회원의 게임 평점을 수집하여 사용한다. 확장성을 해결하기 위해 top-N 규칙이 아닌 평가항목이 많이 겹치는 규칙을 적용함으로써 기존 방법의 정확성을 해치지 않으면서 비교대상의 범위를 줄이고자 한다.

### 2. 협력필터링을 이용한 게임추천

#### 2.1 전체 시스템 구조



[그림 1] 시스템 운용도

위 그림은 제안하는 시스템의 운용도를 보인다. 시스템은 웹 인터페이스로 구축된다. 사용자는 웹브라우저로 회원가입을 하면서 접해본 게임에 대한 평점을 회원정보에 저장한다. 시스템에서는 게임 전문 사이트에서 미리 수집된 게임에 대한 대량의 평점 데이터를 데이터베이스에 저장해 두고, 저장된 자료와 회원정보를 활용하여 협력 필터링을 통해 사용자에게 추천 게임을 제시한다.

본 시스템은 Angular를 사용하여 웹 페이지를 구축하고, Node.js를 이용하여 DataBase에서 회원, 게임에 대한 평점 등 웹 브라우저에서 요청하는 데이터 삽입, 조회, 수정, 삭제 등을 수행한다. Angular는 Google에서 주도해 오픈소스로 개발 중인 프레임워크이며 마이크로소프트웨어에서 개발한 TypeScript를 사용한다. Node.js는 Chrome V8 JavaScript 엔진 위에서 동작하는 이벤트 처리 I/O 프레임워크이며 JavaScript기반이다.

## 2.2 초기 데이터 수집

초기 데이터 수집을 위해 외국 게임 전문 사이트인 Metacritic의 평점 데이터를 크롤링하였다. 수집에서는 Jsoup을 이용하여, 사용자 ID와 각 사용자가 다양한 게임에 대해 내린 평점을 수집한다. Metacritic에 등록되어 있는 PC게임만을 대상으로 하였으며 게임의 항목은 2만여 개, 사용자는 9만여 명, 게임에 대한 평점은 21만여 개를 수집하였다.

사용자의 유사도를 측정하기 위해서는 어느 정도 게임에 대한 평점이 있어야 한다. [1]에서는 20개 이상 평점 정보가 입력되어 있어야 보다 정확한 유사도 측정이 가능하다고 분석했다. 이러한 쿨드 스타트의 문제를 해결하기 위해, 본 시스템에서는 사용자가 신규 가입할 때 다양한 게임에 대한 평가 데이터를 입력하도록 유도하는 사용자 인터페이스를 활용한다.

## 2.3 유사도 측정

협력 필터링을 사용하여 대상 게임에 대한 사용자 평점을 예측하기 위해서는 사용자가 다른 게임에 입력한 평점을 이용하여 측정된 유사도 수치를 사용하여야 한다.

사용자간 유사도를 구하는 대표적인 2가지 방법으로 피어슨 상관관계 공식을 이용한 유사도 측정 방법[2]와 보완 코사인 유사도 측정 방법[3]이 있다. 두 방법은 잘못된 유사도 평가를 초래하는 평가 스케일의 차이를 극복하기 위해 개성을 완화한다는 점이 유사하나, 피어슨 상관관계 공식은 사용자가 평가한 평균으로 개성을 완화하고, 보완 코사인 유사도 측정 방법은 항목에 대한 평균으로 개성을 완화하는 차이점이 있다. 본 논문에서는 두 방법 중 보완 코사인 유사도 측정 방법을 선택하여 사용하였다.

$$\text{sim}(k, l) = \frac{\sum_{i=1}^m (k_i - \bar{r}_i)(l_i - \bar{r}_i)}{\sqrt{\sum_{i=1}^m (k_i - \bar{r}_i)^2} \sqrt{\sum_{i=1}^m (l_i - \bar{r}_i)^2}} \quad (1)$$

식 (1)은 사용자 k와 사용자 l 간의 유사도  $\text{sim}(k, l)$ 을 계산하기 위한 수식을 보인다. 식에서  $k_i$ 는 항목 i에 대한 k의 점수,  $l_i$ 는 항목 i에 대한 l의 점수,  $\bar{r}_i$ 는 항목 i의 모든 사용자의 평균 평가치를 뜻한다.

## 2.4 협력 예측

사용자 k의 항목 i에 대한 예측을 구하기 위해서 GroupLens[2]가 제안한 식 (2)를 사용한다.

$$P_{k,i} = \bar{R}_k + \frac{\sum_{l=1}^n (R_{l,i} - \bar{R}_l) \times \text{sim}(k, l)}{\sum_{l=1}^n |\text{sim}(k, l)|} \quad (2)$$

$P_{k,i}$ 는 항목 i에 대한 사용자 k의 평가 예측이다.  $\bar{R}_k$ 는 사용자 k의 평균 점수를, n은 k와 함께 유사도 공식에 사용될 이웃 l의 수를,  $R_{l,i}$ 는 항목 i에 대한 사용자 l의 평가점수이며  $\bar{R}_l$ 은 사용자 l의 평균 점수를 나타낸다. 이 식에서는 보완 코사인과 같은 원리로 개성을 완화하여 식 (1)에서 얻은 사용자 k와 l의 유사도 점수인  $\text{sim}(k, l)$ 을 사용한다.

[1]에서는 항목에 대한 사용자의 평가 예측을 top-N 규칙을 이용하여 이웃들의 평균값으로부터 편차의 가중치 평균을 수행함으로써 계산하였지만, 본 논문에서는 top-N 규칙을 사용하지 않고 평가한 항목이 많이 중복되는 순서로 이웃을 뽑았다. top-N 방식이 유사도를 높은 이웃을 계산하는 처리시간이 오래 걸려 웹 서비스 상용화에 부합하지 않은 것에 비해, 수정된 방식은 빠른 처리 속도를 보일 수 있다. 또한 top-N 방식은 한 항목만 평가한 사용자와의 유사도가 높을 경우 그 사용자가 이웃으로 지정되어 정확도가 크게 감소하는 문제가 있는 것에 비해, 수정된 방식에서는 평가한 항목이 많이 중복되는 순서로 이웃을 지정하여 보다 높은 정확도를 나타내도록 개선하였다.

## 3. 테스트 및 분석

실험 및 평가는 정확성 테스트와 수행시간 테스트 두 단계로 진행했다. 테스트의 대상으로 쿨드 스타트 문제 해결을 위해 넣어진 평점 데이터 중 40개 이상의 평가를 수행한 사람 20명을 랜덤으로 선발하였다.

테스트 데이터로는 Metacritic을 크롤링한 결과인 게임 정보 2만여 개, 사용자 정보 9만여 개, 평점 정보 21만 개를 사용하였다.

### 3.1 정확성 테스트

정확성 테스트에서는 테스트 대상의 평가 데이터 40개 중 20개를 제외한 나머지 데이터를 제거하고, 제거한 20개 항목에 대한 예상 평점을 시스템으로 도출해낸다. 기존의 평점과 예상 평점이  $\pm 0.5$ 점 이내에 들면 성공으로 보고, 정확도를 측정하여 평균을 계산했다. 이 때, 비교 대상 수는 30명으로 한다.

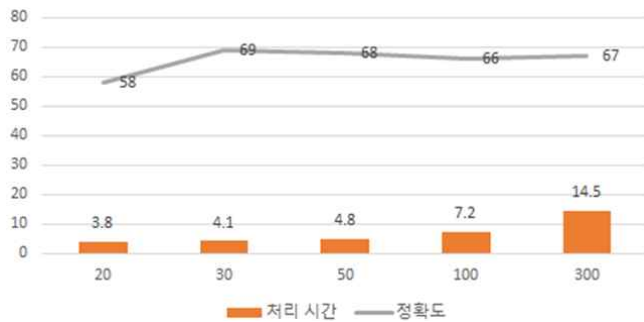
정확성 테스트를 수행한 20명에 대해 평균적으로 61.2%의 정확도를 보였고, 최저 47.16%, 최고 71.7%의 정확도를 보였다. 결과 분석에서는 많은 사용자들이 하지 않은 게임에 대해 평가를 했거나 대다수의 사용자들이 내린 평가와 상반되는 평가를 많이 내린 사용자에게 대해 낮은 정확도를 보였으며, 그 반대의 경우 높은 정확

도를 나타냈다.

### 3.2 수행시간 테스트

수행시간 테스트는 예상 평점을 계산하기 위한 비교 대상 수를 조절 하고 예상 평점 계산을 수행 하면서, 예상 평점 계산 시간을 측정하였다. 이 때, 비교 대상 수를 20명, 30명, 50명, 100명, 300명 순으로 증가하며 테스트를 수행하였다. 그림 2는 결과를 보인다. 그림에서 x축은 비교 대상 수를 보이며, 실선은 정확도(%), 막대그래프는 처리 속도(초)를 보인다.

20명에 대한 테스트 결과를 평균 낸 결과에서 20명에서 처리시간이 3.8초로 제일 짧았지만, 30명을 비교 대상으로 두었을 때 20명과 수행시간에 비해 0.3초 증가하였으나 정확도는 11% 상승하였다. 50명, 100명, 300명과 비교 하였을 때, 정확도 면에서 크게 차이를 보이지 않았으나 수행시간은 많이 증가하였다. 아래의 그림은 비교 대상 수에 대한 정확도와 처리시간 추이를 보인다. 결과에서 협력 필터링은 데이터가 증가함에 따라 처리 속도가 급격히 떨어지는 문제가 존재함을 확인할 수 있었다. 하지만, 데이터가 일정 수준 있지 않으면 정확도가 크게 떨어지기 때문에 성능과 정확도를 조율하는 것이 중요함을 알 수 있다.



[그림 2] 비교 대상 수에 따른 처리시간 및 정확도 추이

## 4. 결론

실험을 통하여 협력 필터링을 이용한 게임 추천 서비스 구현 시 데이터가 증가함에 따라 처리 속도가 낮아짐을 확인하였다. 본 논문에서는 이러한 문제를 해결하기 위해 정확도를 크게 해치지 않으면서 비교 대상의 수를 줄일 수 있는 방법을 제시하였다.

제안 방법은 기존의 방법에 비해 처리 속도 측면에서는 많은 이득이 있지만, 정확도면에서 손해를 야기하기 때문에, 사용자에게 곧 바로 보여줘야 하는 웹 서비스 등에서는 본 논문에서 제시한 방법이 사용자에게 더 나은 서비스를 제공할 수 있을 것으로 보인다.

향후 내용기반 방법을 협력 필터링 방식과 결합하고, 둘의 가중치를 적절히 조절 한다면, 더 나은 처리 속도와 더 나은 정확도를 보일 수 있을 것으로 보인다. 앞으로 내용 기반 방법과 협력 필터링 방식의 조합에 대한 가중치를 적절히 찾아낼 수 있는 방법에 대한 추가적인 연구를 할 계획이다.

## 참 고 문 헌

- [1] 김병만, 이경, 김시관, 임은기, 김주연, “추천시스템을 위한 내용기반 필터링과 협력필터링의 새로운 결합 기법”, 정보과학회논문지 : 소프트웨어 및 응용, 31권, 3호, 332-342, 2004
- [2] Resnick, P., Iacovou, N., Suchak, M., Bergstorm, P. and Riedl, J., “GroupLens: An open architecture for collaborative filtering of Netnews,” Proc. of ACM Conf. on Computer-Supported Cooperative Work, pp.175-186, 1994.
- [3] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, “Item-based Collaborative Filtering Recommendation Algorithms,” Proc. of WWW 10, 2001.