

개인 리뷰를 이용한 영화추천 시스템

김민정*, 박두순**

*순천향대학교 전산학과

**순천향대학교 컴퓨터소프트웨어공학과

e-mail:kmj6906@hanmail.net

A Movie Recommendation System using Individual Review and Meta Data

Min-Jeong Kim*, Doo-Soon Park**

*Dept of Computer Science, SoonchungHyang University

**Dept of Computer Software Engineering, SoonchungHyang University

요 약

최근 많은 추천 시스템들이 연구 되고 있으며, 사용자들에게 의사결정을 도와주는 추천시스템에 대한 중요도가 급증하고 있다. 기존의 영화 추천시스템에서는 희박성의 문제가 제기된다. 본 논문에서는 이러한 문제를 보완하고자 사용자가 영화에 대해 남긴 리뷰로부터 영화키워드를 분석하고 분석된 키워드로부터 가중치를 활용한다. 즉 사용자들로부터 영화에 대한 리뷰를 수집하고 리뷰로부터 각 영화 키워드를 분석해 키워드별 가중치를 활용해 이를 기반으로 영화를 추천한다. 그 결과 사용자에게 만족 할만한 정보를 제공해 효율성을 높이고, 영화에 대한 개인 리뷰를 반영한 영화추천 시스템을 설계 및 구현해 사용자에게 적절한 영화를 추천한다.

1. 서론

최근 스마트기기의 대중화와 웹 2.0의 급속한 발전과 함께 시공의 제약 없이 온라인 서비스를 실시간으로 이용할 수 있게 되었다. 이를 통해 많은 사용자들이 멀티미디어 콘텐츠를 공급하고 공유하는 것이 가능해졌다. 그러나 사용자들이 공급하는 수많은 멀티미디어 콘텐츠들로 인해 사용자는 자신에게 적합한 멀티미디어 콘텐츠를 찾는 데 어려움을 겪고 있다[1]. 그 중에서도 특히 영화에 대한 관심이 급증하는 추세이다. 또한 스마트폰을 가지고 생활하는 시간이 급격히 많아짐에 따라 스마트폰을 활용하여 영화를 예매하고 관람할 수 있는 다양한 방법이 생기고 있다. 따라서 스마트폰을 통해 사람들은 개인의 기호에 따른 맞춤형 영화추천 서비스를 제공받기를 원한다[2]. 개인 맞춤형 영화 추천서비스는 협업 필터링 방법을 통해 제공될 수 있지만 서비스를 주기적으로 받지 않거나, 서비스를 사용하지 않은 사람들에게는 희박성 문제를 가지고 있다. 대부분의 영화 추천 사이트에서는 이러한 문제를 해결하기 위해 선호 장르, 나이, 성별, 최신 개봉작 중심으로 추천한다[3,4].

하지만 추천 시스템의 유용성에도 불구하고 국내에서 영화 추천 서비스를 제공하는 업체는 극히 드물고, 기존의 알고리즘에 대한 한계로 인해 다양한 문제점들이 드러나고 있다. 따라서 이 연구에서는 국내 영화 콘텐츠를 포함하고, 다수의 사용자들에게 만족도를 이끌어 낼 수 있는 추천시스템 방법을 제안하고자 한다.

본 논문에서는 사용자들의 정보를 분석하기 위해 영화

에 대한 개인 리뷰를 이용해 메타데이터인 영화 키워드를 추출하고 각 키워드에 가중치를 활용한 방법을 이용해 새로운 영화를 추천하는 방법을 제공한다. 이러한 방법을 통해 사용자들에게 더 원하는 영화 추천서비스를 제공할 수 있도록 구현하였다.

2. 관련 연구

2-1 추천 시스템

추천 시스템은 과거의 사용자 선호 정보를 분석하여 사용자가 만족할 만한 새로운 아이템을 예측하고, 그 중에서도 높은 만족도가 예상되는 아이템을 추천해 준다.

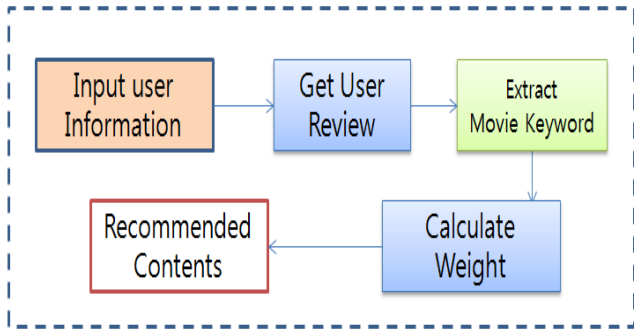
현재까지도 추천 시스템에 관한 연구는 무수히 많이 진행되고 있다. 뿐만 아니라 전자 상거래에서도 제품의 판매 촉진을 위하여 추천 시스템을 도입하고 있다. 대표적으로 Amazon.com, IMDb.com 등을 들 수 있다.

추천시스템의 방법은 크게 콘텐츠 기반 추천 방법(Content-Based Recommendation), 협업적 추천 방법(Collaborative Filtering), 이 둘을 혼합한 혼합형 추천 방법(Hybrid Recommendation)으로 나뉜다. 콘텐츠 추천 방법은 사용자가 과거에 좋아했던 콘텐츠와 유사한 콘텐츠를 추천해 주는 방법이다. 이러한 내용기반 접근은 정보검색(Information retrieval)과 정보 필터링(Information filtering) 연구에 뿌리를 두고 있다[5]. 협업적 추천 방법은 과거에 같은 선호도나 취향을 가지고 있는 사용자들이 좋아했던 콘텐츠를 추천하는 방법을 말한다. 또한, 협업

적 추천 방법은 아이템 추천 방법과 사용자 기반 추천 방법으로 나뉜다[6].

3. 제안하는 방법

본 논문에서 제안하는 방법이자 구현한 추천 방법은 (그림 1)과 같다.



(그림 1) 제안하는 추천 방법

사용자는 회원가입을 통해 개인화 요인 데이터베이스를 생성하고, 사용자가 남긴 영화에 대한 리뷰를 사용해 메타데이터인 영화 키워드를 추출해 분석을 하며 추출된 영화 키워드들은 각각 가중치를 부여해 새로운 추천영화 Top-N 리스트를 생성한다. 영화 키워드에 대한 데이터베이스는 관리자가 XML을 이용해 정의한다. 새로운 사용자는 중요 추천 요소인 영화에 대한 리뷰 데이터베이스가 없어 추천을 할 수 없는 경우, 초기 회원 가입을 통한 생년월일, 성별, 영화 타입 등의 개인화 요인을 받아 사용자와 유사한 사람들을 군집화하고, 군집 내에 속하는 사용자들의 리뷰를 이용해 추천 리스트를 생성한다.

3-1 데이터 수집

본 연구에서 사용한 사용자들의 리뷰와 평점은 <http://www.daum.net>에서 수집한다.

(그림 2)는 Daum 영화 사용자들의 리뷰의 일부를 보여주며 (그림 3)은 10점 만점의 Daum 영화 사용자들의 평점을 보여준다. 본 연구에서는 네티즌 평점을 이용한다.

「우리 할아버지, 아버지의 삶의 이야기」
 이 영화는 한국전쟁부터 현재까지 근현대사를 담은 영화이다. 우리 할아버지와 아버지의 이야기라서 미분들에게 그 시절을 회상하며 볼 수 있고, 젊은 세대들은 그때 그 시절 우리 부모님이 어떻게 살아오셨구나 라는 생각을 하며 볼 수 있다.

(그림 2) Daum Movie의 영화에 대한 사용자들의 리뷰의 일부



(그림 3) Daum Movie의 영화에 대한 평점

3-2 리뷰에서 영화키워드 추출

보통 영화 검색에서 사용하는 메타데이터는 장르, 감독, 배우, 키워드 등이 있다. 이러한 메타데이터들은 각 메타데이터에 해당하는 영화를 대표하는 정보를 가지고 있다. 이 중 리뷰로부터는 메타데이터 중 영화 키워드를 사용하며 영화 키워드 추출을 위해 XML을 이용해 키워드를 정의하고 이를 기반으로 영화를 검색한다. 영화 키워드는 ‘가상현실’, ‘가족’, ‘감동’, ‘사랑’ 등 186개로 나뉘며 각 키워드는 가중치를 가지게 된다. 가중치 부여는 senti 워드넷을 이용해 극성(Polarity)을 분석한다. 극성은 긍정(Positive), 중립(Neutral), 부정(Negative)으로 나타낸다. (그림 4)는 ‘good’의 극성 정도를 나타낸 것이다. 이 단어는 긍정(0.75), 중립(0.25), 부정(0)의 정도를 가지고 있다. 본 논문에서는 중립과 부정의 극성 정도는 이용하지 않고 긍정의 극성만 이용해 가중치를 부여한다.



(그림 4) senti워드넷을 이용한 단어의 극성 정도를 분석한 예

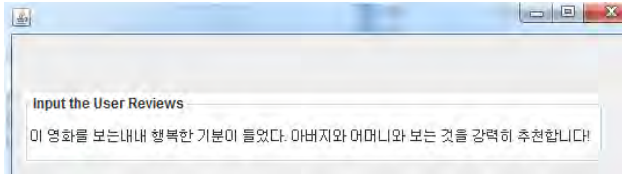
4. 개인 리뷰를 이용한 영화 추천 시스템 구현

제안하는 추천 시스템을 사용하기 위해 사용자의 영화 리뷰를 필수로 하고 입력한 정보는 회원 DB에 저장된다.

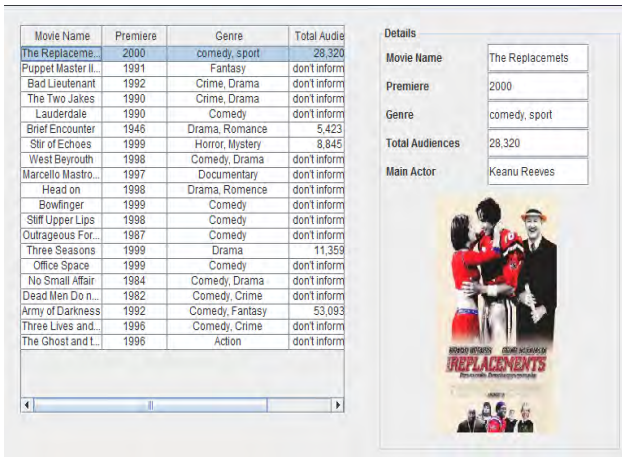
(그림 5)는 DB에 저장된 사용자의 영화별 평점을 나타내며 (그림 6)은 DB에 저장된 사용자의 영화 리뷰를 보여준다. (그림 7)은 추천 시스템 중 가장 많이 사용되는 협업 필터링 방법으로 사용자의 평점을 바탕으로 생성된 영화 20개 리스트를 보여주며 (그림 8)은 개인 리뷰와 가중치를 부여한 영화 키워드를 이용해 새롭게 생성된 영화 20개 리스트를 보여준다.



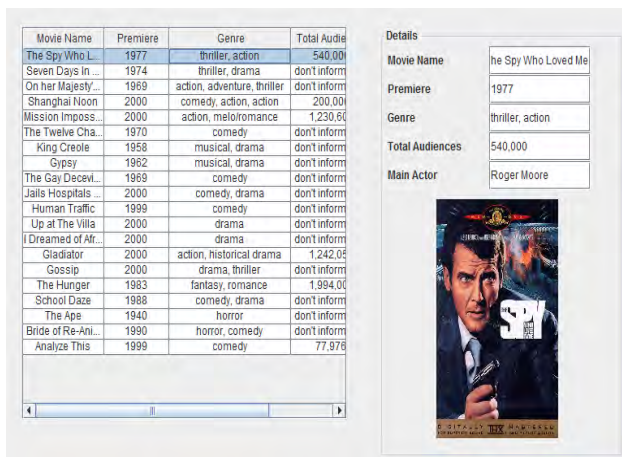
(그림 5) 영화 목록별 사용자 평점 DB



(그림 6) 사용자 리뷰 DB



(그림 7) 협업필터링 방법을 이용한 추천 결과



(그림 8) 개인 리뷰와 영화키워드 가중치를 이용한 추천 결과

본 논문에서 사용하는 추천 시스템에서 가장 중요한 요인은 사용자 개인 리뷰에 속한 메타데이터인 영화 키워드

이다. 이 키워드들의 차이를 두기 위해 각 가중치는 센티워드넷의 긍정 극성값만을 이용해 각각 다른 가중치를 부여하였다.

(그림 8)이 이 단어들의 가중치를 통해 추천한 결과를 보여준다. 그 결과 협업 필터링 방법만 이용한 (그림 7)과 다른 새롭게 추천된 영화 리스트를 나타낸다.

5. 결론

기술의 발달과 휴대기기의 보급은 정보의 폭발적인 증가를 가져왔다. 늘어난 정보는 사용자가 원하는 정보를 빠르고 쉽게 찾는데 어려움을 주고 있어 이를 해결하는 다양한 여러 방법이 제공되고 있지만 개인의 성향을 반영하지 못하기에 직접적인 의사결정에 큰 영향을 주지 못하고 있다. 이러한 개인의 성향을 반영하여 의사결정에 도움을 주는 추천 시스템이 주목을 받고 있다. 추천 시스템에서 주로 사용하는 기법으로 협업 필터링 방법이 많이 사용되고 있다. 하지만 협업 필터링 방법은 충분한 평가치 존재하지 않을 때 추천의 만족도가 떨어지는 희박성 문제가 있다. 희박성 문제를 보완하고자 본 논문에서는 영화에 대한 개인 리뷰에서 메타데이터인 영화 키워드를 추출해 키워드 차이를 두기 위한 다른 가중치를 부여하여 구현한 결과 기존의 개인화 추천 연구에 비해 추천 효과의 성능 향상을 보였다. 향후 연구로는 영화 키워드의 분류기준을 확장하여 정확한 연구와 사용자들의 유동성 있는 리뷰 분석을 통하여 친화적이고 정확한 개인화 추천을 제공받기 위한 영화를 추천하는 시스템으로 연구하고자 한다.

Acknowledgments

This research was supported by Basic Science Research program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF-2014R1A1A4A01007190).

참고문헌

- [1] P. C. George Lekakos, "A hybrid approach for movie recommendation," Multimedia Tools and Applications, vol. 36, pp.55-70, 2008.
- [2] 김수철 "스마트 홈 환경에서 신뢰할 수 있는 사용자 인증 및 추천시스템에 관한 연구" 중앙대학교 학위논문 (박사)
- [3] C.W. Leung, S.C. Chanm and F.-I. Chung, "Integrating collaborative filtering and sentiment analysis : A rating ingerence approach," in Proceedings of The ECAI 2006 Workshop on Recommender Systems, 2006, pp.62-66.
- [4] C.-C. Musat, Y. Liang, and B. Faltings, "Recommendation using textual opinions, " in

Proceedings of the Twenty-Third international joint conference on Artificial Intelligence, 2013, pp.2684-2690.

[5] 천성권 “희박한 평점 데이터를 가진 사용자를 위한 멘토기반의 영화추천 시스템” 서울대학교 학위논문(석사)

[6] P.Wang and H. Ye, “A personalized recommendation algorithm combining slope one scheme and user based collaborative filtering,” in Industrial and Information System, 2009. IIS'09. International Conference on, 2009, pp.152-154.