

개인 리뷰를 통한 영화추천 시스템

김채린*, 박주현*, 두추월*, 박두순**
 *순천향대학교 컴퓨터소프트웨어공학과
 e-mail : kcr0701@sch.ac.kr

A Movie Recommendation System using Individual Review

Chae-Reen Kim, Joo-Hyun Park, Du-qiuyue, Doo-Soon Park
 Dept. of Computer Software Engineering, Soonchunghyang University

요 약

최근 들어, 많은 추천시스템들이 연구 되고 있으며, 대부분은 개인 맞춤형 추천 시스템이 연구되고 있다. 기존의 영화추천시스템에서는 희박성의 문제가 제기된다. 본 논문에서는 희박성에 대해 보완하고자, 개인리뷰에 대한 가중치를 활용한다. 그 결과 사용자에게 정보의 제공에 대해 효율성을 높이고, 사용자마다 영화에 대한 리뷰에 따른 감정 및 사용자의 정보들을 반영한 영화추천시스템을 설계 및 구현한다.

1. 서론

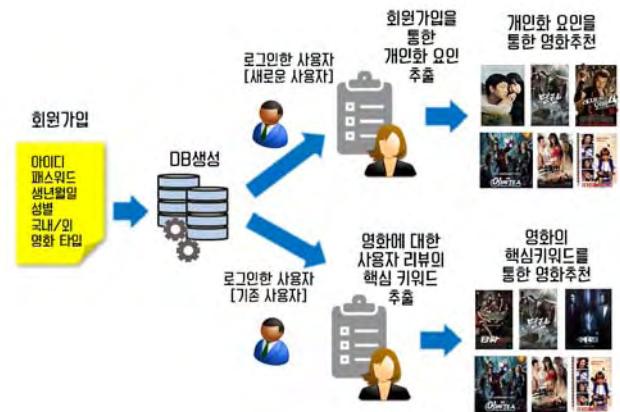
최근 학교 또는 회사가 주 5일제로 단축됨으로 인해 여가생활에 대한 관심이 급증하고 있다[1]. 그 중에서도 특히 영화에 대한 관심이 급증하는 추세이다. 또한 스마트폰을 가지고 생활하는 시간이 급격히 많아짐에 따라 스마트폰을 활용하여 영화를 예매하고 관람할 수 있는 다양한 방법이 생기고 있다. 따라서 스마트폰을 통해 사람들은 개인의 기호에 따른 맞춤형 영화추천 서비스를 제공받기를 원한다[2]. 그러나 개인 맞춤형 영화 추천 서비스는 협업 필터링 방법을 통해서 제공 될 수 있지만 서비스를 주기적으로 도움을 받지 않거나, 몇 번 사용하지 않은 사람들에게는 희박성 문제를 가지고 있다. 대부분의 영화 추천 사이트에서는 희박성 문제를 해결하기 위해서 선호 장르, 나이, 성별 마지막으로 최신 개봉작 중심으로 추천한다.

본 논문에서는 사용자가 직접 작성한 영화리뷰를 바탕으로 관련된 어휘를 추출하고, 추출한 핵심 어휘를 바탕으로 사용자에게 새로운 영화를 추천하는 방법을 제공한다. 또한 입력받은 생년월일, 성별, 국내/외 등의 기호 정보를 이용하여 희박성 문제를 해결한다. 이러한 방법을 통해, 사용자들이 원하는 영화 추천 서비스를 보다 효율적으로 이용할 수 있도록 구현하였다.

2. 개인성향과 오피니언 마이닝 및 형태소분석기를 기반으로 한 추천절차

※ 이 논문은 2014년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. NRF-2014R1A1A4A01007190)

본 논문에서 구현한 추천시스템은 (그림 1)과 같다.



(그림 1) 추천 절차 구성도

기존 사용자는 회원가입을 통한 아이디, 생년월일, 성별, 국내/외, 영화 타입에 대한 개인화요인 데이터베이스를 생성하고, 사용자가 남긴 영화에 대한 리뷰를 통해 형태소 분석기를 이용하여 핵심어휘를 추출해 낸다. 관리자는 영화장르들마다 정해진 핵심어휘에 대해서 작성 후 DB를 생성한다. 영화장르들의 핵심어휘 데이터베이스는 (그림 2)와 같다.

영화장르에 대한 핵심 어휘 데이터베이스는 영화장르 마다의 대표적으로 많이 나오는 사용자의 리뷰를 통해서 관리자가 입력을 해 놓음을 알 수 있다.

로맨스	액션	코미디	스릴러
아름답다	몰입	배꼽	실화
사랑	재밌다	꿀잼	긴장감
반하다	스트레스	판타스틱	잔인함
따뜻하다	박진감	웃음	오싹

(그림 2) 영화장르별 핵심어휘 데이터베이스

그 결과 작성된 영화 장르에 대한 핵심 어휘 데이터베이스와 사용자의 영화 리뷰에 대한 핵심 어휘를 통해서 영화를 추천해 주는 절차를 갖는다.

새로운 사용자는 중요 추천 요소인 리뷰에 대한 데이터베이스가 없어 영화 추천을 할 수 없기 때문에, 초기 회원가입을 통해서 관리자는 새로운 사용자들에 대한 아이디, 생년월일, 성별, 국내/외, 영화 타입인 개인화 요인을 받아 새로운 사용자와 유사한 사람들을 군집화하고, 군집 내에 속하는 사용자들의 리뷰를 이용한 추천리스트를 보여준다. 새로운 사용자에 대한 개인화요인 데이터베이스는 (그림 3)과 같다.

ID	전화번호	생년월일	성별	국내/외
20074256	스릴러	890111	남	국내

(그림 3) 새로운 사용자의 개인화 요인

본 논문은 개인의 성향 중 개인의 특징을 나타낼 수 있는 선호 장르, 나이, 국내/외, 성별을 이용하고 오피니언 마이닝(Opinion Mining)과 형태소 분석기를 통해서 영화 리뷰에 대한 핵심어휘를 추출하여 사용자에게 가장 적합한 영화를 추천하는 추천시스템을 사용한다. 오피니언 마이닝(Opinion Mining)은 다양한 온라인 뉴스와 소셜 미디어의 코멘트 등 콘텐츠들에서 표현된 의견을 추출하고 분류하고 이해하는 컴퓨팅 기술로, 온라인 텍스트 속의 감성과 감동, 주관 및 여러 가지 감정상태를 식별하기 위해 감성분석이 사용되는 방법[3]이고, 형태소분석기는 자연언어를 처리하는 기본적인 도구로, 사용자가 입력한 문장이나 음성 등의 데이터를 분석하여 원하는 형태로 출력 또는 추출하게 된다[4].

3. 영화 리뷰를 통한 영화추천 시스템의 구현

개인 맞춤 추천시스템을 사용하기 위해서는 사용자의 회원가입을 필수로 하고, 회원가입 부분에서 개인화요인으로 사용할 생년월일, 성별, 국내 및 해외, 선호영화장르를 필수적으로 입력받아 회원가입이 이루어진다. 회원가입 양식은 (그림 4)과 같다. 입력한 정보는 회원관리 테이블에 저장되며, 회원관리 테이블은 총 5행으로 이루어져, 선호

도, 국내/외, 생년월일, 성별, ID로 순으로 저장되어, 저장된 순서는 기준치가 높은 것부터 데이터베이스에 저장된다. 회원관리 데이터베이스는 (그림 5)와 같다. 회원관리 데이터베이스에 저장된 기존회원들의 영화에 대한 리뷰의 간략한 데이터베이스의 목록은 (그림 6)과 같다.

아이디	<input type="text"/>
패스워드	<input type="password"/>
패스워드 확인	<input type="password"/>
이름	<input type="text"/>
닉네임	<input type="text"/>
E-mail	<input type="text"/>
생년월일	<input type="text"/>
성별	<input type="text"/>
국내/외	<input type="text"/>
영화 선호도 유형	<input type="text"/>

(그림 4) 회원가입 양식

전화번호	국내/외	생년월일	성별	핵심단어1	핵심단어2	ID
코미디	국내	900503	남	배꼽	스트레스	20094303
코미디	국외	950604	여	웃음	배꼽	20140033
유아	국외	940726	여	귀엽다	유치하다	20134040
액션	국내	910531	남	재밌다	박진감	20104040
스릴러	국내	900423	남	실화	긴장감	20094256
스릴러	국외	920516	남	긴장감	잔인함	20114003
스릴러	국내	940701	여	잔인함	오싹	20134047
로맨스	국내	940729	여	반하다	사랑	20134039

(그림 5) 회원관리 데이터베이스

ID	전화번호	리뷰내용
20134047	스릴러	실화를 바탕으로 만들어서 그런지, 스릴있네요
20114003	스릴러	긴장감 있는 스릴러영화이지만, 실화여서 더 오싹했던 영화
20134039	로맨스	따뜻하고 아름다운 사랑이 하고싶은 영화

(그림 6) 기존 회원들에 대한 영화리뷰 데이터베이스

본 논문에서는 초기 회원가입을 통한 사용자의 생년월일(5%), 성별(5%), 국내/외(10%), 영화 선호도 유형(30%)에 점수를 부여하고, 사용자가 영화 리뷰에 대한 핵심단어(50%)를 이용해 영화를 추천하게 된다. 사용자의 영화리뷰에 대한 핵심단어에 대한 데이터베이스는 (그림 7)와 같다.

ID	전화번호	핵심단어1	핵심단어2
20134047	스릴러	잔인함	오싹
20094303	코미디	배꼽	스트레스
20104040	액션	재밌다	박진감
20114003	스릴러	긴장감	잔인함

(그림 7) 사용자 리뷰에 대한 핵심단어

본 논문에서 사용하는 추천 시스템에서 가장 중요한 요인이라고 할 수 있는 것은 사용자에 따른 영화리뷰에 속한 핵심어휘 이다. 따라서 영화 리뷰 점수를 50%로 많이 할당하고 영화 선호도 유형 점수는 30%, 마지막으로 나이, 성별, 국내/외로 10%의 가중치를 주었다. 본 논문에서는 추천받기 위한 표본 집단을 실화, 긴장, 오싹함이라는 핵심어휘와 20대, 남성, 스릴러에 대한 기준을 통해 추천받기 버튼을 누르게 되었을 때, 핵심단어들을 기준으로 영화장르를 통해서, 그 중 사용자의 초기 회원가입을 통한 개인화요인의 가중치를 통해서 영화를 추천한 결과를 보였다. 그 결과, (그림 8)은 사용자의 영화리뷰에 따른 핵심어휘 및 다른 개인화 요인들을 통해서 추천된 영화 리스트를 나타낸다.



(그림 8) 추천 결과

(그림 8)에서 영화를 선택하여 누르게 되면, (그림 9)과 같은 화면이 출력되면서, 영화에 대한 정보와 영화에 대한 줄거리 및 관련키워드에 대한 정보를 알려줌을 통해서, 사용자의 영화의 리뷰에 대한 키워드를 중심으로 오피니언 마이닝과 형태소분석기를 이용하여 영화 목록을 추천 받아 보여주는 화면이다.

숨바꼭질

- 영화명 : 숨바꼭질 II 107분
- 장르 : 스릴러
- 감독 : 허정
- 출연 : 손현주(성수), 문정희(주희), 전미선(민지)
- 등급 : [국내]15세 관람가

줄거리

사라진 형, 숨바꼭질 압호. 서로 다른 두 가족에게 찾아온 충격적 진실. 가족을 지키기 위한 두 가문의 숨바꼭질 사투가 시작된다!

관련 키워드 및 리뷰

- 관련 키워드
오싹 긴장감, 실화
- 추천 리뷰
오싹하고 소름끼칠만큼 손현주와 문정희 사이의 긴장감은 대단하다. 그리고, 실화라는 점에서 무척이나 놀라웠다.

(그림 9) 추천 영화의 정보

반면에, 중요 추천 요소인 리뷰가 없는 새로운 사용자는 회원가입을 통한 개인화 요인인 성별, 생년월일, 국내/외, 영화타입을 받아 새로운 사용자와 유사한 사람들을 군집화 하고, 군집 내에 속하는 사용자들의 리뷰를 통한 추천 리스트를 보여준다. (그림 3)에 대한 유사한 사용자들의 군집화 된 데이터베이스는 (그림 10)와 같다.

번호	성호도	국내/외	생년월일	성별	핵심단어1	핵심단어2	ID
1	스릴러	국내	900423	남	실화	긴장감	20094256
2	스릴러	국외	920516	남	긴장감	잔인함	20114003
3	스릴러	국내	940701	여	잔인함	오싹	20134047

(그림 10) 유사한 사용자의 군집화 데이터베이스

그 결과, (그림 3)에 대한 새로운 사용자에 대한 영화 추천결과는 (그림 11)과 같다.



(그림 11) 새로운 사용자에 대한 추천 결과

(그림 11)을 통해서 추천된 영화를 누르게 되면 (그림 12)와 같은 화면이 출력되며, 영화에 대한 정보와, 줄거리 관련키워드 및 리뷰에 대한 정보를 알려줌으로써 새로운 사용자에게도 영화를 추천해 줄 수 있도록 한다.

더 테러 라이브

- 영화명 : 더 테러 라이브 II 97분
- 장르 : 스릴러, 드라마
- 감독 : 김병우
- 출연 : 하정우, 이경영, 전혜진
- 등급 : [국내]15세 관람가

줄거리

한강 폭탄 테러의 생생한 충격이 라디오를 통해 독점생생하게 되는데...

관련 키워드 및 리뷰

- 관련 키워드
긴장, 박진감
- 추천 리뷰
하정우 혼자서 이끌어 가는 영화지만, 긴장을 놓치지 않고, 탄탄한 스토리로 이끌어 나간다. 하지만 결말이 아쉬움이 남는다.

(그림 12) 새로운 사용자에 대한 영화추천

4. 결론

본 논문에서는 기존의 개인화 성향을 사용하는 방식에 추가적으로 영화리뷰의 핵심 어휘를 추출하여 해당어휘를 바탕으로 영화를 추천해 주는 기능을 추가함으로써 기존의 개인화 추천연구에 비해 추천 효과의 성능 향상을 보였다. 하지만, 희박성의 문제로 새로운 사용자에게는 개인화 요인을 통해 그와 관련된 기존의 사용자와의 군집화를 통해 영화를 추천해주는 것을 보였다.

향후 연구 과제로는 영화에 대한 리뷰에는 유동성이 있기 때문에, 관리자가 영화에 대한 대표키워드를 지정해주는 것이 아닌, 영화에 대한 대표키워드를 정확히 제시하여 영화를 추천해 준다면 정확하고, 친화적인 시스템으로 구현을 필요로 한다. 또한, 형태소 분석기를 사용 시 조금 더 뚜렷한 명사를 추출 할 수 있도록 하는 것을 연구과제에 핵심을 두고 있다. 예를 들어, 사용자의 영화리뷰에 “이 영화는 판타스틱하고, 나의 생각도 자유롭게 해준 영화이다.”라고 저장되어 있다면, 본 논문에서 현재 구현한 것으로는 문장에서의 명사 및 동사만을 추출하도록 되어 있다. 위의 예시 문장을 통해 본 논문은 판타스틱”만 추출하도록 되어있지만, “자유롭게”란 형용사도 같이 핵심 어휘로 추출하는 것을 활용하는 것을 필요로 한다.

더불어, 다른 요소들인 상영년도, 영화사 또는 주연 등을 활용하여 영화를 추천해주는 시스템 연구가 필요 할 것으로 보이며, 더불어 조건에 부모 일 때, 아이들의 연령에 따라 함께 볼 수 있는 영화에 대한 보폭을 넓혀 정보를 입력 할 수 있고, 그러한 데이터에 따른 영화를 추천해주는 페이지를 추가적으로 만들어서 더 나은 연구를 하게 될 것이다. 이를 통해, 필수 요소들에 대해서 관리하는 것에 대한 구현을 추가적으로 생성함으로써 발전된 연구과제로 구현 할 것이다. 그 결과, 구체적인 시스템 구현으로 정확한 영화 추천하는 시스템으로 확장 할 수 있을 것으로 예상된다.

참고문헌

- [1] 김소현 “주 5일제 수업 도입에 따른 청소년들의 여가 시간 활용 현황” 한양대학교 학위논문(석사)
- [2] 김수철 “스마트 홈 환경에서 신뢰할 수 있는 사용자 인증 및 추천시스템에 관한 연구” 중앙대학교 학위논문(박사)
- [3] 네이버 지식백과, IT용어사전 ‘오피니언 마이닝’
- [4] 이증영, 신병훈, 이공주, 김지은, 안상규 “COM기반의 다목적 형태소 분석기를 이요한 명사 추출기” 한국정보과학회언어공학연구회 99 제11회 한글 및 한국어 정보처리 학술대회 및 제1회 형태소 분석기 및 품사태거 평가 워크숍