

동시출현단어분석을 이용한 연관영화정보 분석 연구

An Analysis of Related Movie Information Using The Co-Word Method

최상희 (Sanghee Choi)*

초 록

최근 이용자들이 정보를 공동생산하고 소비하는 웹기반 서비스들이 활발해지면서 이용자가 정보를 이용한 기록이나 이용자가 습득한 정보를 활용하여 생산한 다양한 부가 정보들이 다시 이용자에게 제공되고 있다. 또한 쌍방향으로 이용자들이 소통할 수 있는 정보채널이 다양해짐으로써 공통된 관심사를 가진 이용자의 정보소비 경험을 공유할 수 있는 방법이 활발하게 모색되고 있다. 이 연구에서는 동시출현정보 분석기법과 자이중심 네트워크 분석 기법을 적용하여 IMDB 서비스의 기존 이용자들이 자신이 보고 싶거나 좋아하는 영화를 선별하여 만들어 놓은 영화리스트에 나타난 정보를 토대로 특정 영화를 좋아하는 이용자가 선호할 만한 다른 영화를 찾아낼 수 있도록 연관영화정보를 다각적으로 표현하였다. 한 영화를 기준으로 연관 영화, 감독, 장르로 분석을 한 결과 영화의 테마나 주인공성향과 같은 다양한 자질로도 연관영화가 연결되었고 감독의 경우 영화내용보다는 감독의 인지도에 영향을 받는 것으로 나타났다. 또한 영화는 주제의 복잡성이 큰 것으로 나타나 장르가 연관영화정보를 제공하기에 적합하지 않은 것으로 분석되었다.

ABSTRACT

Recently, many information services allow users to collaborate to produce and use information. Sharing information is also important for users who have similar taste or interest. As various channels are available for users to share their experiences and knowledge, users' data have also been accumulated within the information services. This study collected movie lists made by users of IMDB service. Co-word analysis and ego-centered network analysis were adapted to discover relevant information for users who chose a specific movie. Three factors of movies including movie title, director and genre were used to present related movie information. Movie title is an effective feature to present related movies with various aspects such as theme or characters and the popularity of directors affects on identifying related directors. Genre is not useful to find related movies due to the complexity in the topic of a movie.

키워드: 동시출현단어, 용어계량분석, 네트워크분석, 이용자추천정보, 자이중심 네트워크
informetric analysis, co-word analysis, network analysis, user recommendation,
ego-centered network

* 대구가톨릭대학교 도서관학과 조교수(shchoi@cu.ac.kr)

■ 논문접수일자: 2014년 11월 20일 ■ 최초심사일자: 2014년 12월 1일 ■ 게재확정일자: 2014년 12월 17일
■ 정보관리학회지, 31(4), 161-178, 2014. [http://dx.doi.org/10.3743/KOSIM.2014.31.4.161]

1. 서론

데이터가 대용량으로 축적되면서 이용자들이 원하는 정보를 획득하는데 겪는 어려움을 해소하고자 하는 것은 정보학을 연구하는 많은 연구자들이 관심을 갖는 기본 과제이다. 특히, 이용자들이 원하는 정보를 스스로 찾아내는 것이 점차 달성하기 힘든 문제가 되어 가면서 유사한 상황에 처한 다른 이용자들의 의견이나 사례를 활용하는 것이 대안으로 활발히 연구되고 있다. 이러한 방안 중 하나가 다수의 이용자들이 특화된 정보검색 시스템이나 전자상거래 시스템, 소셜네트워크 등에서 정보활용 과정을 공유하고 서로 협력하면서 축적한 정보를 기반으로 새로운 이용자에게 연관정보를 제공하는 것이다. 특히 정보의 주제나 장르가 구체적이지 않고 정보의 소비행태가 예측되기 어렵다면 기존에 구축되어 있는 정보체계나 지식구조로는 유사정보를 추천해주는 과정은 쉽지 않을 것이다. 예를 들어 학술정보를 소비하는 경우는 연관된 정보를 추적하는 기준이 주제로 한정되는 경우가 많지만 일반 생활정보나 오락정보를 소비하는 경우 이용자의 취향 등이 결합되어 연관된 정보를 제공하게 되는 기준이 다양해진다.

이 연구에서는 이용자의 취향을 예측하는데 다양한 기준이 적용될 수 있는 유형의 정보로 영화를 선택하여 영화를 소비하는 이용자들이 생산한 정보를 활용하여 연관 영화정보를 제공할 때 활용할 수 있는 방안을 제시하고자 한다. 이와 같은 목적으로 인터넷에서 가장 활성화되어 있는 영화 사이트인 IMDB에서 이용자들이 다른 이용자와 공유하기 위해 만든 공개된 영화

위시리스트(wish list)를 수집하였다. 영화 위시리스트는 이용자의 취향이 반영된 영화목록으로서 같이 보면 좋거나 같이 봐서 좋았던 영화, 보고 싶은 영화 등을 선정하여 생성한 이용자 생산 정보이다. 실제 영화를 보러갈 때 영화 전문가가 추천한 영화보다는 자신과 비슷한 취향을 가진 주변인들이 추천한 영화가 더 좋을 수 있다는 가정 하에 자신이 선택한 특정 영화를 '보고 싶은' 또는 '보고 좋았던' 영화로 선정한 이용자들의 영화 위시리스트를 활용하는 방안이다. 자신과 같은 영화를 좋아하는 이용자의 영화리스트 속에 포함된 다른 영화에 대한 정보를 추출하여 연관영화를 선택할 때 활용할 수 있는 정보로 제공하고자 하는 것이다.

이 연구에서는 다양한 이용자가 영화 위시리스트에 선택한 영화정보를 분석하여 특정 영화의 연관정보를 다각적으로 분석하고자 하였다. 특정 영화의 연관정보를 추출하는데 적용한 기법은 White의 자아(Ego)중심 인용분석기법이다(White, 2000). 자아중심 인용분석기법 중 하나로 특정 연구자를 자아로 규정한 후 이 연구자와 동시인용된 저자들을 분석하여 이 연구자의 이미지를 분석하는 인용이미지 방식이 있다. 이 연구에서는 영화 위시리스트에 특정영화와 동시인용된 영화간 관계를 분석하는데 인용 이미지방식을 적용하였다.

이용자가 선별한 영화목록 리스트는 특정 영화를 기준으로 이 영화가 포함된 영화목록으로 검색하여 위시리스트 100건을 데이터로 수집하였다. 기준으로 사용된 영화는 'Life of Pi'이며 수집된 영화 위시리스트를 대상으로 영화제목, 감독, 장르별 동시출현정보를 통계적으로 분석하였다. 또한 연관 영화정보를 추출하기

위하여 동시출현빈도로 분석된 결과를 토대로 하여 빈도수 5 이상인 영화제목, 감독, 장르를 대상으로 네트워크 분석을 한 후 시각화하였다. 네트워크 분석에 적용된 기법은 클러스터링 기반 네트워크(Clustering_based Network)(이재윤, 2007)인데 클러스터링 알고리즘이 적용된 네트워크 기법을 적용한 이유는 'Life of Pi'를 중심으로 수집된 정보를 군집화하여 각 군집의 특성을 분석한 후 이를 통해 영화의 이미지를 다각적으로 표현하고자 함이다. 영화의 이미지를 표현하기 위해서 도출된 영화 군집은 해당 영화와 연관된 영화정보로 활용될 수 있을 것이다. 네트워크내 노드간 유사도는 코사인 계수로 산출되었으며 계층적 클러스터링 기법은 ward 기법을 적용하였다. 생성된 네트워크는 NodeXL 프로그램으로 시각화하였으며 네트워크 분석을 통해 생성된 군집을 대표할 수 있는 데이터를 선정하기 위하여 최근접이웃중심성 척도(Nearest Neighbor Centrality: NNC)(이재윤, 2013)와 군집내 출현빈도를 활용하였다.

2. 선행연구

2.1 동시출현정보 분석연구

동시출현정보를 분석하려는 시도는 계량서지적 연구분야에서 동시인용을 대상으로 다양하게 수행되어 왔는데(Åström, 2007; Chen, 1999; White and McCain, 1998) 동시인용을 중심으로 연구주제를 파악하거나 연구동향을 분석하던 연구는 점차 인용이 아닌 텍스트 정보를 활용하는 영역으로 확장되었다. 1980년대에 Callon

(Callon et al., 1983; Callon, Law, & Rip, 1986)은 동시출현단어 분석기법을 소개하였는데 이 기법은 색인에 사용된 용어를 이용하여 지적구조를 분석하는데 적용되었다. 이 연구에서 제시한 동시출현단어 분석 결과는 인용기반 분석이나 주제분류보다 더 객관적으로 문서의 주제 유사도를 측정해내는 것으로 평가받았다(Morris & Van Der Veer Martens, 2008). 초창기 동시출현단어 분석은 디스크립터, 주제명표목 등 통제된 어휘를 사용한 경우가 일반적이었다(Callon et al., 1983; McCain, 1995). 그러나 이후 제목이나 초록 등에 나타난 자연어를 대상으로 동시출현분석을 수행한 연구가 다양한 계량연구에서 나타났다(Buter & Nuyons, 2002; Porter, 2005; Janssens et al., 2006; Seo & Yu, 2013; 김하진, 송민, 2014). 동시출현단어 분석에 통제어와 자연어가 모두 사용되면서 동시출현단어 분석에 통제어휘와 비통제어휘를 사용할 때 발생하는 각각의 문제점을 해결하기 위하여 통제어휘 디스크립터를 동시 출현하는 자연어로 프로파일링 하는 디스크립터 프로파일링 방안이 제시되기도 했다(김관준, 이재윤, 2007).

동시출현 정보분석에 네트워크 기법을 적용하여 학술 연구동향 및 주제 영역을 분석하는 연구도 다양하게 시도되고 있다(White, 2000; Bar-Ilan, 2008). 최근 국내에서도 동시출현단어 분석에 네트워크 기법을 적용한 연구가 활발하게 수행되어 한국어 교육학의 지적구조를 파악하거나(강범일, 박지홍, 2013) 오픈엑세스 분야의 연구주제 영역을 파악한 시도가 있다(서선경, 정은경, 2013). 또한 독서에 관련된 학술지 논문을 수집해서 동시출현한 색인어를 분석한 후 독서분야의 연구가 이론적 연구에서 실

증적으로 변화하고 있는 것을 파악한 사례가 있다(장령령, 홍현진, 2014).

동시인용정보를 활용하여 특정 주제분야의 연구주제나 연구동향을 파악하는 목적 외에도 특정 연구자가 연구하는 주제에 대한 분석에 동시출현정보와 네트워크 기법을 적용하는 시도가 있었는데, 가장 널리 알려진 연구로서는 White의 저자 인용이미지 연구가 있다(White, 2000). 이 연구에서는 사회 네트워크 분석에서 제시한 연구기법은 자아 중심 네트워크 분석(ego-centered network analysis)을 적용하여 연구자의 연구 주제를 파악하고자 하였다. White는 인용정체성, 인용이미지, 인용이미지구축자, 공저자 등 4가지 방법을 제시하였는데 이 연구에서 제시한 분석방법 중 후속연구에서 많이 활용된 것은 인용정체성과 인용이미지이다. 연구자나 인물 단위의 분석이었던 인용정체성과 인용이미지 분석기법은 저널 단위로도 확장되어 Journal of Documentation이라는 복합주제를 다루는 학술지를 대상으로 수행된 사례가 있다. 이 연구에서는 인용정체성과 인용이미지 기법을 적용하여 학술지의 주제영역을 표현하였다(Nebelong-Bonnevie & Frandsen, 2006). 2011년 수행된 연구에서도 정보학 분야의 학술지 3종을 대상으로 학술지의 주제영역을 이와 유사한 방법으로 분석하려는 시도가 있었다(Tray, 2011).

국내연구에서 인용정체성 기법을 적용한 연구로서는 정보학 분야의 특정 연구자를 대상으로 이 연구자가 인용한 저자집단의 동시출현빈도분석을 통하여 연구주제를 분석한 사례가 있다(이재윤, 2012). 이 연구에서는 이 연구자가 인용한 주요 저자의 동시인용을 네트워크 기법과 클러스터링 기법을 적용하여 분석한 결과로

13개의 저자 군집을 산출하여 각 군집별 주제를 분석함으로써 특정 저자의 연구 주제를 표현하고자 하였다. 인용이미지 분석기법이 적용된 국내 사례로서는 음원서비스 분야에 적용하여 특정 가수의 이미지를 동시출현정보 분석을 통해 표현하여 이를 음원추천에 적용하려는 연구가 있다(최상희, 2014). 이 연구에서는 음원을 서비스하는 사이트 이용자들이 생산한 공개앨범을 활용하였다. 특정 가수를 선정하여 이 가수가 포함된 공개앨범을 수집한 후 이 가수와 동시출현한 곡들의 가수를 분석하여 특성가수의 이미지를 표현하였다. 또한 동시출현한 곡의 통계분석을 통하여 해당 가수를 선택한 이용자가 선호할 만한 연관 곡을 추천하였다.

2.2 정보추천 시스템

연관정보를 제공하기 위해서는 제공할 만한 정보를 추려내는 것이 중요한 과정이다. 이러한 연구는 정보추천시스템 분야에서 활발하게 수행되고 있는데 정보추천 시스템은 1990년대에 정보서비스 분야에 소개되었다(Manure, Parted, & Robinson, 2000). 정보추천시스템(Recommender System)은 일반적으로 3단계에 걸쳐 추천정보를 생성한다. 첫 번째 단계는 이용자들의 입력 데이터에서 선호도 정보를 획득하는 것이며 두 번째 단계는 획득한 정보를 기반으로 추천정보를 생성하는 것이며 마지막 단계는 이를 이용자에게 표현하는 것이다(Kant & Bharadwaj, 2013). 이와 같이 이용자에게 정보를 추천하는 방식은 협력적 필터링 접근방식과 내용기반 접근방식, 지식기반 접근방식, 혼합방식으로 분류된다(Carrer-Neto et al., 2012; Adomavicius

& Tuzhilin, 2005; Burke, 2000).

협력적 필터링 접근방식은 이용자의 취향을 분석하여 비슷한 취향을 가진 다른 이용자들의 소비행태를 분석한 후 이를 기반으로 연관 아이템을 추천하는 방식이다. 즉, 이용자를 기준으로 하여 특정 이용자가 속할 수 있는 이용자 집단을 분석해내서 해당 이용자 집단이 평가한 정보를 활용하는 것이다. 내용기반 방식은 이용자가 선호하는 아이템을 분석하여 아이템을 기준으로 연관정보를 발견해내는 것이다. 이용자가 선택한 아이템과 유사한 특징을 가졌거나 또는 이 아이템과 유사하게 평가받은 다른 아이템을 연관 아이템으로 제안하는 방식이다. 따라서 이 방식에서는 이용자의 피드백을 받는 것이 매우 중요한 과정이다. 지식기반 접근방식은 이용자의 선호도와 기존 아이템간의 관계를 분석하여 이용자를 프로파일링 하는 방식이며 혼합 방식은 위 세 방식 중 두 가지 이상의 방식을 혼용하여 추천정보를 도출해내는 것이다.

최근에는 협력적 필터링을 하는데 적용해야 하는 요소를 다양화하는 시도로서 최신성에 대한 활용성을 분석한 연구가 수행되었는데 그 결과 내용이나 이용자 취향 외에도 시간적 요소를 협력적 필터링에 적용할 경우 성능이 향상되는 것으로 나타났다(Zhang, Peng, & Phelan, 2014).

영화분야에서도 다양한 측면에서 연관영화 정보를 제공하는 영화추천 시스템에 대한 연구가 수행되고 있다. Web 2.0기반 소셜 웹 정보를 활용한 연구(Carrer-Neto et al., 2012)와 엔트로피 이론을 이용자 프로파일링에 적용하여 이용자의 영화선호 다양도를 분석한 연구(Ni et al., 2014)가 발표되었는데 이러한 연구들은 영화추천을 하는데 복합적 요소를 적용하

고자 하는 시도로 평가된다. 또한 정보추천을 하는 접근방식을 혼합하여 적용한 연구로는 내용 분석과 이용자 프로파일링을 혼합하여 영화를 추천하는 방식을 제안한 사례(Pera & Ng, 2013)가 있으며 유사한 국내사례로는 협력적 필터링과 내용기반 접근 방식을 혼합하여 영화추천을 시도한 경우가 있다(김상화, 오병화, 김문중, 양지훈, 2012).

연관영화정보를 제공하기 위하여 이 연구에서는 특정 영화를 기준으로 유사한 이용자 집단이 선정한 정보를 선별하여 활용하는 방식을 적용했으므로 협력적 필터링 기법을 기반으로 하고 있다. 그러나 선별한 정보 내에서 정보간 내용 유사성과 패턴을 분석한다는 측면에서는 내용기반 방식도 적용하였다고 할 수 있다. 따라서 이 연구에서는 복합적 협력추천 방식을 적용하였다고 할 수 있다.

3. 연구방법 및 연구데이터

영화추천을 목적으로 이 연구에서 활용한 데이터는 이용자들이 여러 영화를 선택하여 다른 이용자와 공유하기 위해 만든 공개된 영화위시리스트이다. 영화 위시리스트는 인터넷에서 가장 많은 영화정보를 제공하고 공유하는 IMDB에서 수집하였다. 영화 위시리스트는 IMDB의 이용자가 특정 테마나 장르, 주제 등 다양한 기준으로 같이 보면 좋을 만한 영화를 선정한 영화 목록이다.

이 연구에서는 특정 영화를 기준으로 이 영화를 선택한 이용자가 같이 보면 좋을 만한 영화를 다각적으로 추천하는 것을 목적으로 영

화 위시리스트에서 동시출현단어를 분석하고 연관정보를 분석하기 위하여 영화 위시리스트를 특정 영화기준으로 수집하였다.

데이터 수집 기준으로 선정한 영화는 'Life of Pi'로 영화 'Life of Pi'가 포함된 영화 위시리스트 100개를 수집하여 각 위시리스트 내에 동시출현한 영화와 동시출현한 영화의 감독 및 장르를 분석하였다. 이 영화를 선정한 이유는 이 영화의 장르가 복합적이며 영화의 내용과 주제가 다각적인 측면을 가지고 있어 연관된 영화를 추천하기 위해서는 이 영화를 소비하는 이용자들의 정보를 기반으로 한 협력추천의 필요성이 대두되는 대표적인 사례라고 판단되었기 때문이다.

수집한 위시리스트 100개에서 'Life of Pi'와 동시출현한 영화는 총 3,220개이며 중복되는 영화를 처리한 결과 1,491개 영화가 산출되었다. 영화정보는 영화제목, 감독과 장르로 분할하여 1차 통계분석을 하였다. 통계분석 결과 1,491개 영화를 감독한 영화감독은 총 944명으로 조사되었고 영화 장르는 복합장르로서 409개가 있는 것으로 분석되었다. IMDB에서 영화장르를 복합장르를 허용하기 때문에 하나의 영화에 장르 디스크립터가 1-6개까지 할당되어 있는 것으로 나타났다. 이 연구에서는 복합장르와 단일장르를 모두 고려하여 복합된 형태로도 통계분석하였고 복합장르로 할당된 모든 단일장르를 별도로 추출

하여 단일장르별로도 분석하였다.

이 연구에서는 학술문헌 연구에서 적용된 White의 자아(Ego)중심 인용분석기법(White, 2000)을 적용하여 영화이미지 분석을 하였다. 학술문헌 중심으로 수행되어온 인용이미지의 동시인용분석 단위를 영화 이미지분석에 적용한 사항은 <표 1>의 자아중심 이미지 분석을 위한 동시출현정보 분석 단위 비교와 같다. 분석단위는 영화제목과 감독, 장르인데 영화를 브라우징 하거나 영화에 대한 추천을 받을 때 영화 제목 외에도 감독과 장르는 보편적으로 활용되는 요소이다. 따라서 이 연구에서는 특정 영화를 중심으로 연관정보를 도출해내기 위한 단위로 영화제목, 감독, 장르를 선정하였다.

영화 위시리스트에서 분석된 동시출현정보는 1차적으로 통계 분석하여 빈도별 순위를 산출하였다. 또한 빈도 5 이상 동시출현한 데이터는 별도로 추출하여 네트워크 분석을 하였다. 네트워크 분석은 영화 'Life of Pi'와 동시출현한 영화, 감독, 장르간의 관계를 조사하기 위하여 적용되었으며 연관된 정보를 범주화하기 위하여 클러스터링 기법이 결합된 네트워크 분석방법을 선택하였다. 즉, 'Life of Pi'를 중심으로 수집된 연관정보의 군집관계를 표현하여 영화의 이미지를 다각적으로 표현하고자 함이다. 네트워크 분석에 적용된 기법은 CBnet인데 CBnet은 클러스터링 기반 네트워크(Clustering_based Network)

<표 1> 자아중심 이미지 분석을 위한 동시출현정보 분석 단위 비교

구분	학술저자 - 인용이미지	영화제목 - 영화이미지
데이터 단위	학술저자가 포함된 문헌의 참고문헌 목록	영화제목이 포함된 영화위시리스트의 영화정보목록
분석 단위	저자	영화제목, 감독, 장르
분석결과로 도출된 연관정보	특정 저자와 함께 인용되는 연구자	특정 영화와 같이 인용되는 영화, 감독, 장르

로 ward기법과 같은 계층적 클러스터링 알고리즘을 이용하여 네트워크를 구축하는 기법이다(이재운, 2007). CBnet은 노드간 유사성을 측정하여 네트워크의 서브그래프를 설정하는데 계층적으로 클러스터링함으로써 네트워크내에서 군집과 관계를 동시에 표현하는 방식이다. 이 연구에서는 CBnet에 적용하는 계층적 클러스터링 기법으로 ward기법을 선택하였고 유사도 계수는 코사인 계수를 사용하였다. CBnet으로 산출된 군집은 9개이며 생성된 CBnet은 NodeXL 프로그램으로 시각화하였다.

계층적 클러스터링 네트워크 분석을 통해 생성된 영화제목과 영화감독 군집을 대표할 수 있는 데이터를 선정하기 위하여 최근접이웃중심성 척도(Nearest Neighbor Centrality: NNC)를 활용하였다. 최근접이웃중심성은 네트워크에서 형성된 클러스터내에서 다른 노드로부터 최근접 이웃으로 선정되는 경우 수를 바탕으로 측정되는 것으로 가중네트워크에서 지역중심성을 측정하는 지표로 사용된다(이재운, 2013). 이 연구에서는 네트워크 분석시 형성된 군집내에서 중심이 되는 노드를 측정하는 기준으로

이 척도를 사용하였고 최근접이웃중심성 값이 가장 높은 데이터를 해당 군집을 대표하는 데이터로 선정하였다.

장르 네트워크 분석을 위해서는 복합장르의 특성을 반영하여 다시 군집내에 속한 복합장르의 단일장르 요소를 통계분석한 후 상위 3개 단일장르 요소를 생성된 군집을 대표할 수 있는 데이터로 선정하였다.

4. 동시출현정보를 기반으로 한 영화정보 협력추천

4.1 연관 영화 분석

'Life of Pi'가 포함된 영화 위시리스트에 동시출현한 영화 1,491개 중 가장 출현빈도가 높은 영화는 <표 2>와 같다. 가장 많이 동시출현한 영화는 'Django Unchained'와 'Gravity'인데 두 영화와 'Life of Pi'의 연관관계를 세부적으로 살펴보면 'Django Unchained'는 2012년 같은 해에 개봉한 영화로 주제적 연관성보다는

<표 2> 동시출현빈도 상위 10개 영화

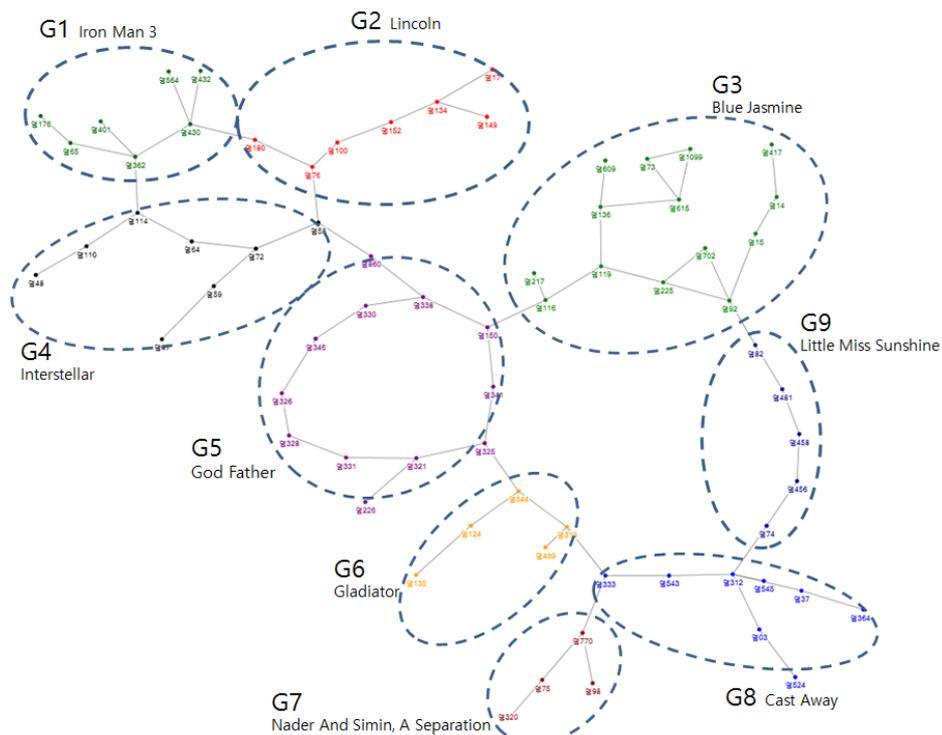
동시출현빈도 순위	Title	동시출현빈도
1	Django Unchained	22
1	Gravity	22
3	Intouchables	18
4	The Dark Knight Rises	17
5	12 Years a Slave	16
5	Cloud Atlas	16
7	Fight Club	15
7	Her	15
7	Inception	15
10	Argo	14

개봉시기로 인해 이용자들이 위시리스트에 같이 선택한 것으로 분석되었다. 반면 'Gravity'는 우주에 고립되어 구조되는 과정을 그린 영화로 바다에서 조난을 당하여 구조될 때 까지 상황을 그린 'Life of Pi'와 내용적으로 유사성을 띄고 있다. 상위 10개 영화를 세부적으로 살펴보았을 때 'Django Unchained'를 포함하여 'The Dark Knight Rises', 'Cloud Atlas', 'Argo' 등 총 4개의 영화가 동일한 개봉시기인 것으로 나타난 이용자가 여러 영화를 선택하는데 있어 개봉시기는 중요한 요소인 것으로 분석되었다.

함께 출현한 주요 영화들을 기반으로 'Life of Pi'와 연관되는 영화정보를 군집화하기 위하여 동시출현빈도가 5번이 넘는 영화를 추출하

여, 영화들 간의 관계를 CBnet으로 표현한 결과는 <그림 1>의 영화 'Life of Pi' - 동시출현 영화 네트워크 분석과 같다.

군집 1은 'Iron Man 3'과 'X-Men: First Class', 'Edge of Tomorrow'와 같은 영화가 모인 군집으로 공상과학 전쟁영화 군집이며 군집 2는 'Lincoln'과 'Les Miserables' 등 역사, 사건이 중심으로 군집화된 것으로 분석되었다. 또한 군집 4는 'Interstellar', 'Guardians of the Galaxy', 'Gravity'가 포함되어 우주가 테마인 것으로 해석되었으며 군집 6은 'Gladiator', 'A beautiful mind', 'Forest Gump' 등 역경을 이겨내는 인물 영화 집합으로 분석되었다. 또한 군집 8에 'Cast away'를 기점으로 'Captain Phillips', '127 Hours'



<그림 1> 영화 'Life of Pi' - 동시출현 영화 네트워크 분석

와 같은 조난영화들이 군집화되었고 군집 7로는 'Nader and Simin: A separation'과 Das Leben der Anderen(타인의 삶) 포함되어 오락정보다는 작품성이 있는 영화가 클러스터링 된 것으로 나타났다. 군집 9는 'Little Miss Sunshine', 'Eternal Sunshine of the Spotless Mind', 'Big Fish' 등 잔잔하면서 환상적 이미지를 보여주는 영화들이 군집화된 것으로 분석되었다. 형성된 군집 중 군집 1과 군집 4는 'Life of Pi'와 주제적 연관성보다는 영화주제의 인기도에 따라 이용자들에게 선택된 군집으로 해석되지만 군집 6이나 군집 8의 경우 내용상 조난을 이겨내는 인물영화라는 측면에서 내용간 유사성을 가지고 있다고 분석되었다. 군집 7은 'Life of Pi'가 단순 오락성 영화가 아닌 작품성이 있는 영화로 평가받는 측면에서 연관된 정보라 할 수 있으며 군집 9의 경우 인간의 내면을 판타지와 결합한 영화들이 속해있는 군집이라고 해석되어 표현적인 측면에서 유사성을 띄고 있는 것으로 조사되었다.

군집 3과 5는 상대적으로 큰 군집임에도 불구하고 특별한 유사성이 없는 영화들로 군집화된 것으로 나타나 일부 영화들은 주제적 연관성이 없이 동시선정되고 있는 것으로 나타났다. 또한 통계적으로 동시출현빈도가 높은 상위 10개의 영화를 분석하였을 때 영향을 미치는 요인으로 나타난 영화 개봉시기는 클러스터링 네트워크 분석에서는 영화를 군집화하는 요인으로는 역할을 하고 있지 못한 것으로 분석되었다.

4.2 연관 감독 분석

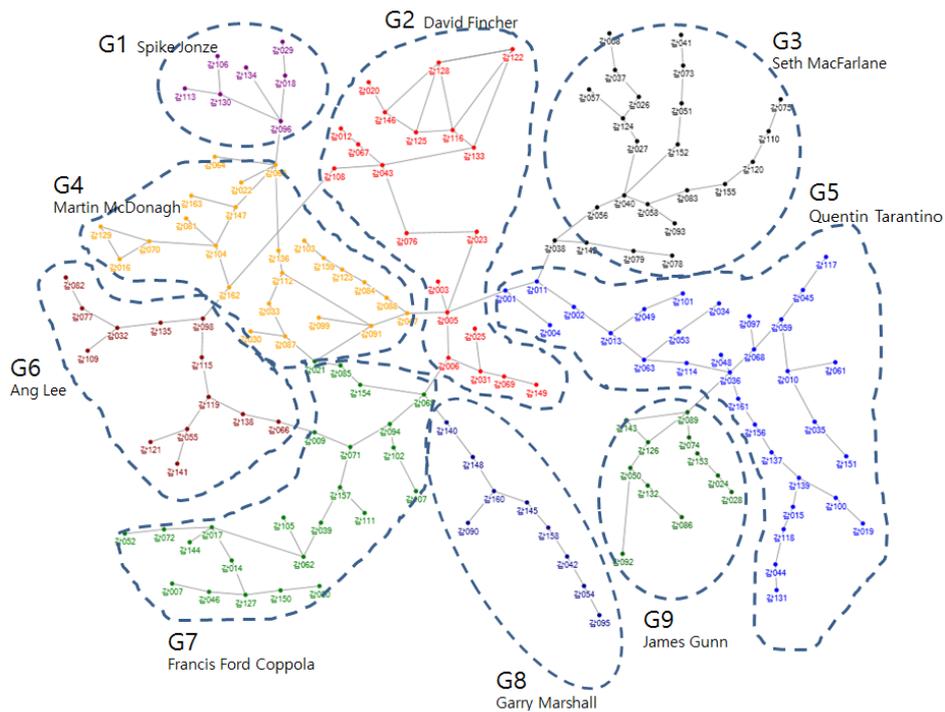
감독은 영화를 선정하는데 중요한 요소이다.

따라서 이 연구에서는 특정 영화에 연관된 정보를 추천하는 요소로 감독을 활용하고자 하였다. 'Life of Pi'가 포함된 영화 위시리스트에 동시출현한 영화를 감독한 사람 중 가장 출현빈도가 높은 감독은 <표 3>의 동시출현빈도 상위 10개 영화감독과 같다. 가장 많이 동시출현한 영화감독은 'Christopher Nolan'으로 'The Dark Knight Rises'와 'Interstella' 등 최근에 인기 있는 영화를 비롯하여 총 7개 영화가 동시출현한 것으로 분석되어 감독의 경우 영화내용의 연관성보다는 감독의 최근 인기도에 영향을 받는 것으로 해석되었다. 이와 같은 현상은 동시출현빈도 순위가 높은 감독들에게서도 공통적으로 나타나는 현상으로 'Life of Pi'의 개봉시거나 주제연관성과 상관없이 한 감독의 영화가 5-8개 정도 선택된 것으로 분석되었다. 이용자들이 감독을 기준으로 연관영화를 선정할 때 감독의 인기도에 영향을 많이 받는 것으로 나타났다. 또한 인기있는 감독의 경우 전작까지 연계해서 한 감독의 영화를 복수로 선정하는 성향이 있는 것으로 조사되었다. 이 같은 현상은 동시출현빈도 수에서도 나타나고 있는데 동시출현한 영화제목의 경우 최대 동시출현빈도수가 22 수준이었으나 감독의 최대 동시출현빈도수는 이의 약 3.5배에 해당하는 71에 해당되어 연관영화의 경우 다양한 영화로 분산되는 성향이 나타나지만 감독의 경우 인기감독으로 집중되는 현상이 있는 것으로 나타났다.

<그림 2>는 영화 'Life of Pi' - 동시출현 영화 감독 네트워크 분석은 위시리스트에 함께 출현한 영화를 감독한 사람 중 동시출현빈도가 5번이 넘는 감독을 추출하여, 감독들 간의 관계를 CBnet으로 표현한 결과이다.

〈표 3〉 동시출현빈도 상위 10개 영화감독

동시출현빈도 순위	감독명	동시출현빈도
1	Christopher Nolan	71
2	Peter Jackson	54
3	Steven Spielberg	51
4	Quentin Tarantino	49
5	David Fincher	47
6	Martin Scorsese	41
7	Robert Zemeckis	36
8	David O. Russell	27
9	Frank Darabont	27
10	Alfonso Cuarón, Marc Webb	26



〈그림 2〉 영화 'Life of Pi' - 동시출현 영화감독 네트워크 분석

각 군집에 소속된 감독간의 유사성을 분석한 결과 감독한 영화의 주제적 연관성보다는 인기 도나 이미지에 영향을 받는 것으로 나타났다. 군집 1은 'Spike Jonze', 'Wes Anderson' 등이

군집화되어 인디영화의 이미지가 있는 감독들이 모인 것으로 분석되었고 군집 3은 'Seth MacFarlane', 'Drew Goddard', 'Stephen Chbosky' 등이 모여 촉망받는 신예감독 집단으로 해석되

었으며, 군집 4는 'Clint Eastwood', 'Florian Henckel von Donnersmarck', 'Sofia Coppola', 'Gus Van Sant' 등 흥행보다는 작품성이 있는 영화를 감독하는 집단으로 조사되었다. 군집 5의 경우는 현재 'Tarantino', 'Nolan' 등 지명도가 높은 할리우드 감독들이 모여 있는 군집으로 나타났고 군집 6은 박찬욱, 봉준호, M. Night Shyamalan 등이 모여 있어 아시아계 할리우드 감독들이 모여진 다국적 군집으로 해석되었다. 군집 7은 'Francis Ford Coppola'를 중심으로 'Roman Polanski', 'Tim Burton', 'Stanley Kubrick', 'Woody Allen', 'James Cameron' 등 각 분야의 거장으로 조명받는 감독들이 모인 군집으로 해석되었다.

4.3 동시출연영화를 기반으로 한 장르 분석

장르는 특정 영화를 찾는 것이 아니라 막연한 요구를 가지고 영화를 찾을 때 또는 특별히 표현할 구체적인 정보요구가 없을 때 영화를 브라우징하는 도구로 사용되는 가장 일반적인

요소이다. 따라서 대부분의 영화사이트에서는 영화를 장르별로 분류하여 이용자에게 제공하고 있으나 실제 장르로 브라우징을 하게 될 경우 한 장르에 영화가 너무 많이 나오거나 영화의 장르를 결정하기 어려운 경우가 많다. IMDB에서도 영화를 장르별로 제공하고 있는데 영화 주제의 복합성 때문에 영화장르에 해당되는 장르를 모두 할당할 수 있도록 허용하는 복합장르 체제로 운영하고 있다. 이 연구에서 이미지 분석대상으로 선정한 영화 'Life of Pi'는 장르가 'Adventure', 'Drama'로 설정되어 있었다.

'Life of Pi'와 연관된 영화를 브라우징할 수 있는 장르를 추천하기 위하여 위시리스트에 동시출현한 영화의 장르를 기반으로 단일 장르요소와 복합 장르로 나누어 분석하였다. 동시출현한 영화의 장르로 나타난 항목을 단일장르 요소로 분할하여 통계적으로 분석한 결과는 <표 4>의 단일장르요소의 동시출현빈도 상위 10위와 같다.

동시출현빈도가 가장 높은 'drama'의 경우 총 403건의 영화에 할당되어 전체 영화의 27%에 해당하는 것으로 나타났다. 다음 순으로는 'action', 'comedy', 'crime', 'adventure' 순으로

<표 4> 단일장르요소의 동시출현빈도 상위10위

동시출현빈도 순위	단일 장르요소	동시출현빈도
1	drama	403
2	action	284
3	comedy	278
4	crime	121
5	adventure	114
6	animation	90
7	biography	80
8	horror	33
9	mystery	24
10	documentary	24

나타나 'Life of Pi'와 같이 선호되는 영화의 주요 장르로 분석되었다.

복합장르로 통계분석을 한 결과는 <표 5>의 복합장르요소의 동시출현빈도 상위 10위와 같은 데 상위 5위까지 장르의 세부 요소를 살펴보면 'drama'와 'comedy', 'romance'가 주요 장르로 나타났다. 이중 'romance'는 단일장르요소로 분석하였을 때는 동시출현빈도가 많지 않았던 항목이었다. 5위 이하에서는 'thriller'가 'romance'와 마찬가지로 세부 장르를 구성하는 주요 요소로 나타나 메인 주제장르로서 연관영화를 결정하는 역할을 하지 못하지만 장르를 구체화하는 역할을 수행하고 있는 것으로 해석되었다.

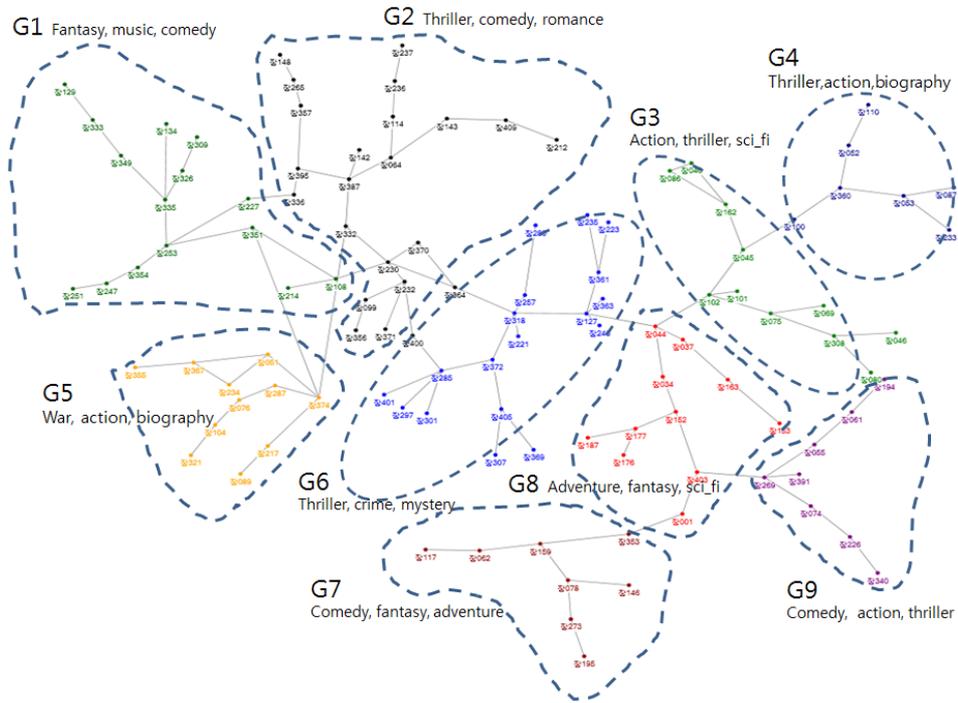
'Life of Pi'의 장르로 선정된 'adventure'와 'drama'와의 관계를 살펴보면 단일장르요소로 분석하였을 경우와 복합장르요소로 분석하였을 경우 모두 'drama'가 가장 동시출현빈도가 높은 장르에 속해있어 다른 장르와 결합해서도 'Life of Pi'의 연관영화를 선택하는 일차적 선호요소로서 역할을 하고 있는 것으로 분석되었다. 그러나 전체적으로 많은 타 장르요소와 결합하여 나타나고 결합된 요소도 다양하게 나타

나 단일장르는 연관영화정보를 제공하는 용도에 적합하지 않은 것으로 조사되었다.

'Life of Pi'와 이용자의 영화 위시리스트에 동시출현한 영화의 장르 중 동시출현빈도가 5번이 넘는 복합장르를 추출하여 복합장르들 간의 관계를 CBnet으로 표현한 결과는 <그림 3>의 영화 'Life of Pi' -동시출현 복합장르 네트워크 분석과 같다. 각 장르군집을 대표하는 키워드는 군집내에 속한 복합장르를 단일장르 요소로 분할한 후 통계분석을 하여 'drama'를 제외하고 상위 3개 단일장르 요소를 생성된 군집을 대표할 수 있는 데이터로 선정한 것이다. 'drama'는 단일장르요소로 가장 많은 타 장르와 결합되어 있었고 통계분석을 하였을 경우 모든 장르군집에서 상위권을 차지하여 장르를 변별하는 능력이 없는 불용키워드에 해당하는 것으로 분석되었다. 따라서 연관된 영화장르 군집을 변별하는 요소로는 적합하지 않아 'drama'를 제외하고 장르요소를 선정하게 되었다. 형성된 장르 군집은 'fantasy, music, comedy', 'thriller, comedy, romance', 'action, thriller, sci_fi', 'thriller, action, biography', 'war, action, biography',

<표 5> 복합장르요소의 동시출현빈도 상위 10위

동시출현빈도 순위	장르명	해당되는 영화 수
1	drama, romance	71
2	comedy, drama	56
3	comedy, drama, romance	48
4	comedy, romance	38
5	crime, drama, thriller	34
6	drama, thriller	25
7	drama, mystery, thriller	24
8	action, adventure, sci_fi	23
8	action, crime, thriller	23
8	crime, drama, mystery, thriller	23



〈그림 3〉 영화 'Life of Pi' -동시출현 복합장르 네트워크 분석

'thriller, crime, mystery', 'comedy, fantasy, adventure', 'adventure, fantasy, sci-fi', 'comedy, action, thriller'로 나타났다.

클러스터링 네트워크 분석결과 새롭게 형성된 복합장르에서는 기존의 장르 분석 상위권에서는 나타나지 않았던 'war'나 'music'같은 요소가 나타났는데 이는 장르를 클러스터링 네트워크로 새롭게 군집화하여 제시하였을 경우 drama와 같은 장르적 변별력이 떨어지는 대분류 장르에 속한 영화집단의 세부 장르를 변별해 낼 수 있는 방안이 될 수 있는 것으로 해석되었다.

5. 결론

이 연구에서는 특정 영화를 기준으로 연관 영화를 추천하는 정보를 영화제목, 감독, 장르별로 산출하기 위하여 동시출현정보 분석 및 자아중심 이미지 분석 기법과 클러스터링 네트워크 기법을 적용하여 아래와 같은 사항을 도출했다.

첫째, 동시출현정보를 통계적으로 분석하여 상위 10개의 항목을 선정하였을 경우 영화제목과 영화감독, 장르별로 각기 다른 특성이 분석되었다. 동시출현빈도가 높은 최상위 영화의 경우 개봉시기가 동일한 인기영화와 내용적인 측면에서 유사한 영화가 모두 포함되어 있는 것

로 분석되었다. 연관영화를 추천할 경우 내용적 유사성도 중요한 요소이지만 개봉시기의 유사성과 개봉당시 해당 영화의 인기도도 영향을 미치는 요소이다. 즉, 복수의 영화를 소비하는데 있어서 내용간 유사성도 영화를 선정하는데 중요한 기준이지만 인기도나 개봉시기의 유사성도 영향을 미치는 요소인 것이다. 이와 같은 현상은 영화 위시리스트의 특성으로 해석될 수 있으며 한 영화를 기점으로 연관 영화를 선택하는 기준이 다양한 것으로 조사되었다.

영화감독을 동시출현빈도 기준으로 분석한 결과 상위로 분석된 감독은 영화내용의 연관성보다는 감독의 최근 인기도에 영향을 받아 선정된 것으로 해석되었다. 동시출현빈도 순위가 높은 감독의 영화는 'Life of Pi'의 개봉시기나 주제연관성과 상관없이 선택되어 한 감독의 영화가 5-8개 정도 포함된 것으로 분석되었다. 이와 같은 현상은 이용자들이 위시 리스트에 연관 영화를 선정할 때 특정 감독의 전작까지 연계해서 복수로 선정하는 성향이 있는 것으로 해석되었다. 그러나 특정 감독을 선정하는 기준이 영화의 내용적 유사성보다 영화감독의 최근 인지도에 따른 것으로 나타났다. 이는 학술정보에서 연관 연구자가 주제적 유사성을 띄는 특성과는 차별화되는 현상으로 감독으로 연관영화를 제공할 때 고려해야 할 사항이다.

장르의 경우 동시출현빈도를 기준으로 분석하였을 경우 'drama'가 출현빈도가 27%에 해당하여 'Life of Pi'를 기준으로 하여 장르로 연관영화를 선택할 경우 'drama'가 가장 중요한 장르인 것으로 조사되었다. 그러나 'drama'의 경우 다양한 타 장르와 결합되어 사용되는 기본 요소로 작용하는 측면이 있어 연관 영화를 구체

적으로 선별해주는 역할을 하고 있지 못하는 것으로 분석되었다.

둘째, 자아중심 이미지 분석기법을 적용하여 'Life of Pi'와 연관된 영화정보를 동시출현한 영화정보로 표현한 결과 영화제목에 따라 형성된 군집 중 일부 군집은 주제적 연관성보다는 영화주제의 인기도에 따라 이용자들에게 선택된 군집으로 해석되었고 일부 군집은 내용간 유사성으로 형성된 것으로 조사되었다. 또한 한 군집은 작품성이 있다고 평가되는 이미지로 비슷한 이미지를 가진 작품성이 있는 영화로 군집화되었으며 영화의 영상 표현적인 측면에서 유사성이 있는 영화끼리 군집화된 경우도 있었다. 즉, 동시출현한 영화를 군집화하였을 경우 영화 내용 외에도 인기도, 작품성 이미지, 영상적 이미지 등 다양한 기준으로 연관 영화정보가 추출될 수 있는 것으로 분석되었고 이렇게 영화의 다각적인 요소로 연관정보가 추출된 것은 향후 영화 추천시스템에서 고려할 만한 사항이다. 그러나 동시출현빈도 분석에서 중요한 요소로 분석된 개봉시기는 클러스터링 네트워크 형성과정에서 연관영화를 군집화하는 기준으로서 역할을 하지 못하는 것으로 나타났다.

영화감독을 군집화하여 네트워크 관계도를 그린 결과 각 군집은 감독한 영화의 주제적 연관성보다는 인기도나 이미지에 영향을 받는 것으로 나타났다. 한 군집은 인디영화의 이미지가 있는 감독들이 모인 것으로 조사되었고 신예감독 집단, 흥행보다는 작품성이 있는 영화를 감독하는 집단, 지명도가 높은 할리우드 감독들이 모여 있는 군집, 다국적 할리우드 감독들이 모여진 군집, 세부 영화 분야의 거장으로 조명받는 감독들이 모인 군집 등 감독의 특성별로 군

집이 형성되었다. 감독의 특성별로 군집화된 정보는 연관 영화를 다양한 측면으로 찾고자 할 때 유용한 연관정보가 될 수 있다.

장르를 군집화한 결과는 연관된 영화 복합장르를 새롭게 생성하여 보여줄 수 있다는 측면에서는 의의가 있었으나 영화의 특성상 장르간 구별이 명확치 않고 5-6개 이상까지 복합장르로 설정되어 있어 단일 장르만으로는 연관 영화를 추천하는데 적절하지 않은 것으로 나타났다. 이는 일반적으로 영화를 브라우징할 때 장르가 가장 보편적으로 사용된다는 사실을 전제로 할 때 추후 장르를 기준으로 영화정보를 추천할 경우 추천 효과성을 고려해서 적용해야 할 사항이다.

일반적으로 연관정보를 제공하려고 할 경우 가장 먼저 고려되는 것은 내용의 유사성이다. 그러나 위와 같이 분석된 결과를 고찰해보면 영화는 내용의 유사성 외에도 영화 개봉시기, 영화의 이미지, 영화의 표현형식, 감독의 인지도, 감독 이미지, 국적 등 다양한 요소로 연관정보가 형성되는 것으로 조사되었다. 정작 내용을 대표하는 장르는 연관 영화정보를 제공하는데 효율적이

지 못한 것으로 나타나 영화처럼 내용이 복잡적이고 표현하기 어려운 정보는 내용간 유사성을 비교하여 연관 정보를 추출하는데 한계가 있는 것으로 분석되었다. 따라서 영화나 소설 같이 주제가 명확하지 않은 비학술적 정보를 대상으로 연관 정보를 제공하기 위해서는 연관 정보를 선택하는데 영향을 미치는 다양한 요소를 분석하여 활용하는 것이 바람직할 것이다.

마지막으로 이 연구를 수행하면서 도출된 한계점은 연관 영화를 추천할 때 특정 영화를 기준으로 하여 위시리스트를 수집할 때 영화제목만을 기준으로 한 것이다. 위 연구분석 결과에서도 나타났듯이 영화와 감독이 군집화되는 현상은 다르므로 연관 영화정보를 수집할 때에도 차별화 하여 적용해야 할 것이다. 즉, 'Life of Pi'를 기점으로 이용자의 영화 위시리스트를 수집할 때 영화제목이 출현한 목록과 이 영화를 감독한 감독명이 출현한 목록을 이중으로 수집하여 분석 비교하면 보다 다각적인 측면으로 연관정보를 산출할 수 있는 것으로 예측되어 후속연구가 필요한 것으로 판단되었다.

참 고 문 헌

- 강범일, 박지홍 (2013). 프로파일링 분석과 동시출현단어분석을 이용한 한국어교육학의 정체성분석. 정보관리학회지, 30(4), 195-213. <http://dx.doi.org/10.3743/KOSIM.2013.30.4.195>
- 김상화, 오병화, 김문중, 양지훈 (2012). 협력적 필터링과 콘텐츠 정보를 결합한 영화 추천 알고리즘. 정보과학회논문지: 소프트웨어 및 응용, 39(4), 261-268.
- 김관준, 이재윤 (2007). 연구 영역 분석을 위한 디스크립터 프로파일링에 관한 연구. 정보관리학회지, 24(4), 285-303. <http://dx.doi.org/10.3743/KOSIM.2007.24.4.285>
- 김하진, 송민 (2014). 동시출현단어 분석을 통한 국내외 정보학 학회지 연구동향 파악. 정보관리학회지, 31(1), 99-118. <http://dx.doi.org/10.3743/KOSIM.2014.31.1.099>

- 서선경, 정은경 (2013). 동시출현단어 분석기반 오픈 액세스 분야 지적구조에 관한 연구. 한국비블리아학회지, 24(1), 207-227. <http://dx.doi.org/10.14699/kbiblia.2013.24.1.207>
- 이재운 (2006). 지적구조 분석을 위한 새로운 클러스터링 기법에 관한 연구. 정보관리학회지, 23(4), 215-231. <http://dx.doi.org/10.3743/KOSIM.2006.23.4.215>
- 이재운 (2007). 국내 광역 과학 지도 생성 연구. 정보관리학회지, 24(3), 363-383. <http://dx.doi.org/10.3743/KOSIM.2007.24.3.363>
- 이재운 (2012). 자기 인용 네트워크와 인용 정체성을 이용한 연구자의 연구 이력 분석에 관한 연구. 정보관리학회지, 29(1), 157-174. <http://dx.doi.org/10.3743/KOSIM.2012.29.1.157>
- 이재운 (2013). tnet과 WNET의 가중 네트워크 중심성 지수 비교 연구. 정보관리학회지, 30(4), 241-264. <http://dx.doi.org/10.3743/KOSIM.2013.30.4.241>
- 장령령, 홍현진 (2014). 학술지 중요도와 키워드 순서를 고려한 단어동시출현 분석을 이용한 독서분야의 지적구조 분석. 한국비블리아학회지, 25(1), 295-318. <http://dx.doi.org/10.14699/kbiblia.2014.25.1.295>
- 최상희 (2014). 이용자 추천정보를 기반으로 한 가수 이미지 분석. 제21회 한국정보관리학회 학술대회 논문집, 7-10.
- Åström, F. (2007). Changes in the LIS research front: Time-sliced cocitation analyses of LIS journal articles, 1990-2004. *Journal of the American Society for Information Science and Technology*, 58(7), 947-957.
- Bar-Ilan, J. (2008). Which h-index? - A comparison of WoS, Scopus and Google Scholar. *Scientometrics*, 74(2), 257-271.
- Buter, R. K., & Noyons, E. C. M. (2002). Using bibliometric maps to visualise term distribution in scientific papers, In *Proceedings of Sixth International Conference on Information Visualisation*, 697.
- Callon, M., Courtial, J.-P., Turner, WA, & Bauin, S. (1983). From translations to problematic networks: An introduction to co-word analysis. *Social Science Information*, 22, 191-235.
- Callon, M., Law, J., & Rip, A. (Eds.) (1986). *Mapping the Dynamics of Science and Technology: Sociology of Science in the Real World*. London, Macmillan Press.
- Carrer-Neto, W., Hernandez-Alcaras, M. L., Valencia-García, R., & García-Sánchez, F. (2012). Social knowledge-based recommender system. Application to the movies domain, *Expert Systems with Applications*, 39(12), 10990-11000.
- Chen, C. (1999). Visualising semantic spaces and author co-citation networks in digital libraries. *Information Processing & Management*, 35(3), 401-420. [http://dx.doi.org/10.1016/S0306-4573\(98\)00068-5](http://dx.doi.org/10.1016/S0306-4573(98)00068-5).

- Janssens, F., Leta, J., Glänzel, W., & Moor, B. (2006). Towards mapping library and information science. *Information Processing & Management*, 42(6), 1614-1642.
- Kant, V., & Bharadwaj, K. K. (2013). Integrating collaborative and reclusive methods for effective recommendations: A fuzzy Bayesian approach. *International Journal of Interlligent Systems*, 28(11), 1099-1123.
- McCain, K. W. (1995). R&D themes in information science: A preliminary co-descriptor analysis, *Proceedings of the 5th Biennial Conference of the International Society for Scientometrics and Informetrics*. Pine Forest, Il. June 7-10, 275-282.
- Nebelong-Bonnevie, E., & Freandsen, T. F. (2006). Journal citation identity and journal citation image: A portrait of the *Journal of Documentation*. *Journal of Documentation*, 62(1), 30-57.
- Ni, J., Zhang, Y., Zhao-Long, H., Song, W., Qiang, G., & Lui, J. (2014). Ceiling effect of online user interests for the movies. *Physica A*, 402, 134-140.
- Pera, M. S., & NG, Y. (2013) A group recommender for movies based on content similarity and popularity. *Information Processing and Management*, 49(3), 673-687.
- Porter, A. L. (2005). QTIP: Quick technology intelligence processes. *Technological Forecasting & Social Change*, 72, 1070-1081.
- Seo, Eun-Gyoung, & Yu, So-Young (2013). Detecting research trends in Korean information science research, 2000-2011. *Journal of the Korean Society for Information Management*, 30(4), 215-239. <http://dx.doi.org/10.3743/KOSIM.2013.30.4.215>
- Todorov, R. (1992). Displaying content of scientific journals: A co-heading analysis. *Scientometrics*, 23(2), 319-334.
- Tsay, M. Y. (2011). A bibliometric analysis and comparison on three information science journals: JASIST, IPM & JOD, 1988-2008. *Scientometrics*, 89(2), 591-606.
- Tseng, Y., Lin, C., & Lin, Y. (2007). Text mining techniques for patent analysis. *Information Processing & Management*, 43(5), 1216-1247.
- White, H. D. (2000). Toward ego-centered citation analysis. In B. Cronin, & H.B. Atkins(Eds.). *The web of knowledge: A festschrift in honor of Eugene Garfield* (pp. 475-496). Medford, New Jersey: Information Today, Inc.
- White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science, 1972-1995. *Journal of the American Society for Information Science*, 49(4), 327-355.
- Zhang, L., Peng, L. F., & Phelan, C. A. (2014). Novel recommendation of user-based collaborative filtering. *Journal of Digital Information Management*, 12(3), 165-175.

• 국문 참고문헌에 대한 영문 표기
(English translation of references written in Korean)

- Choi, S. H. (2014). Analysis of singer's image using user recommended song data, Proceeding of the 21st Annual Conference of Korean Society for Information Management, 7-10.
- Kang, B., & Park, J. (2013). Profiling and co-word analysis of teaching Korean as a foreign language domain. *Journal of the Korean Society for Information Management*, 30(4), 195-213.
- Kim, H. J., & Song, M. (2014). A study on the research trends in domestic/ international information science articles by co-word analysis. *Journal of the Korean Society for Information Management*, 31(1), 99-118.
- Kim, P. J., & Lee, J. Y. (2007). Descriptor profiling for research domain analysis. *Journal of the Korean Society for Information Management*, 24(4), 285-303.
- Kim, S., Oh, B., Kim, M., & Yang, J. (2012). A movie recommendation algorithm combining collaborative filtering and content information. *Journal of KIISE: Software and Application*, 39(4), 261-268.
- Lee, J. Y. (2006). A novel clustering method for examining and analyzing the intellectual structure of a scholarly field. *Journal of the Korean Society for Information Management*, 23(4), 215-231.
- Lee, Jae Yun (2007). Making a science map of Korea. *Journal of the Korean Society for Information Management*, 24(3), 363-383. <http://dx.doi.org/10.3743/KOSIM.2007.24.3.363>
- Lee, J. Y. (2012). Exploring a researcher's personal research history through self-citation network and citation identity. *Journal of the Korean Society for Information Management*, 29(1), 157-174.
- Lee, Jae Yun (2013). A comparison study on the weighted network centrality measures of tnet and WNET. *Journal of the Korean Society for Information Management*, 30(4), 241-264. <http://dx.doi.org/10.3743/KOSIM.2013.30.4.241>
- Seo, S. K., & Chung, E. K. (2013). Domain analysis on the field of open access by co-word analysis. *Journal of Korean Biblia Society for Library and Information Science*, 24(1), 207-227.
- Zhang, L. L., & Hong, H. J. (2014). Examining the intellectual structure of reading studies with co-word analysis based on the importance of journals and sequence of keywords. *Journal of Korean Biblia Society for Library and Information Science*, 25(1), 295-318.