

Freebase 기반의 추천 시스템 시각화

홍명덕*, 하인애*, 조근식**

Visualized recommender system based on Freebase

Myung-Duk Hong*, Inay Ha*, Geun-Sik Jo**

요약

본 논문에서는 영화 추천을 위해 사용자들이 명시적으로 표시한 신뢰 정보를 이용하여 소셜 네트워크와 유사하게 신뢰 네트워크를 생성하고, 그 사용자들의 연결 정도를 이용하여 추천 시스템에 적용하며, 추천 정보는 시각화 방법을 이용하여 제공하는 방법을 제안한다. 이를 통해 사용자가 명시적으로 신뢰 관계를 표현한 신뢰 네트워크에서 숨겨진 신뢰 관계를 추론한다. 시각화된 추천 정보는 영화, 음악, 인물 등 다양한 도메인에 대한 정보를 구조화된 형태로 제공하는 Freebase를 이용하였으며, 시각화 방법은 다음 3가지와 같다. (1) 사용자가 제공받고자 하는 영화의 수만큼 영화 포스터로 시각화하고, (2) 추천된 영화 중 특정 영화를 선택하면 영화 감독, 주연 배우, 장르 등의 부가적인 정보를 시각화하여 제공한다. 마지막으로 (3) 신뢰 기반의 사용자들 중 임의로 몇 명을 이웃 사용자로 선택하여 추천한다. 본 논문에서는 시각화 방법을 적용함으로써 추천 수 또는 이웃 사용자의 수, 그리고 부가 정보 요청 등 사용자의 의견(요구)을 바탕으로 추천하기 때문에 사용자의 의사 결정 능력을 향상시킬 수 있다. 뿐만 아니라 본 논문에서 제안하는 추천 시각화 방법을 통해 동적으로 사용자들의 요구를 반영할 수 있고, Freebase, LinkedMDB, 위키피디아 등 현존하는 LOD의 정보 재사용을 통해 보다 풍부하게 추천 정보를 제공할 수 있다.

▶ Keywords : 추천 시각화, Freebase(LOD), 신뢰 네트워크, 사용자 모델링

Abstract

In this paper, the proposed movie recommender system constructs trust network, which is similar to social network, using user's trust information that users explicitly present. Recommendation on items is performed by using relation degree between users and information of recommended item is provided by a visualization method. We discover the hidden relationships via the constructed trust network. To provide visualized recommendation information, we employ

•제1저자 : 홍명덕 •교신저자 : 조근식

•투고일 : 2013. 8. 8, 심사일 : 2013. 8. 20, 게재확정일 : 2013. 9. 6.

* 인하대학교 컴퓨터정보공학과(Dept. of Computer and Information Engineering, Inha University)

** 인하대학교 컴퓨터정보공학과(School of Computer and Information Engineering, Inha University)

※ 이 논문은 2013년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. 2011-0015484).

Freebase which is large knowledge base supporting information such as movie, music, and people in structured format. We provide three visualization methods as the followings: i) visualization based on movie posters with the number of movies that user required. ii) visualization on extra information such as director, actor and genre and so on when user selected a movie from recommendation list. iii) visualization based on movie posters that is recommended by neighbors who a user selects from trust network. The proposed system considers user's social relations and provides visualization which can reflect user's requirements. Using the visualization methods, user can reach right decision making on items. Furthermore, the proposed system reflects the user's opinion through recommendation visualization methods and can provide rich information to users through LOD(Linked Open Data) Cloud such as Freebase, LinkedMDB and Wikipedia and so on.

▶ Keywords : Visualized recommender system, Freebase(LOD), Trust-network, user modeling

I. 서 론

스마트 시대 또는 빅데이터 시대에 접어들며 블로그, 트위터 등 다양한 형태의 소셜 네트워크 플랫폼이 활성화되고, 이에 따라 소셜 네트워크를 기반으로 한 다양한 정보 및 콘텐츠가 폭발적으로 증가하고 있다. 이 대량의 정보 및 콘텐츠 중에서 사용자가 필요로 하는 정보를 발견하려면 많은 시간과 노력을 필요로 한다. 그래서 사용자들의 행동(behavior)이나 히스토리를 분석하여 사용자들이 선호할 수 있는 정보를 예측하여 제공하는 추천 시스템을 서점, 음악, 영화, 뉴스 등 다양한 도메인의 사이트에서 쉽게 접할 수 있다[1].

추천 시스템은 [2]에 의해 총 3세대로 분류되며, 1세대는 사용자들의 명시적·암묵적 선호도 성향에 따라 규칙 기반의 정보를 처리하는 것으로 정적으로 진행된다. 즉, 사용자가 암묵적이든 명시적이든 특정 아이템이나 콘텐츠를 평가하고, 평가된 이 정보를 기반으로 규칙을 발견한다. 예를 들어, 이 사용자는 '로맨스 영화를 선호 한다' 또는 '추리 소설을 좋아 한다' 등 사용자가 평가한 정보를 통해 규칙을 발견한다. 그리고 이를 기반으로 새로운 아이템이나 콘텐츠, 또는 사용자가 평가하지 않은 아이템이나 콘텐츠를 추천하여 제공한다. 2세대는 사용자들의 명시적·암묵적 선호도 성향에 따라 콘텐츠 기반의 사용자 행동 처리, 협업적인 사용자 행동처리 그리고 하이브리드 사용자 행동을 기반으로 추천하는 것으로 동적으로 처리한다. 사용자가 평가한 아이템이나 콘텐츠의 주요 키워드를 추출하거나(콘텐츠 기반), 다른 사용자들과 공통으로

평가한 아이템이나 콘텐츠를 기반으로 두 사용자간의 유사도를 계산(협업적 기반) 마지막으로 이 두 방법을 결합하는 하이브리드 방법으로 선호도(성향)가 유사한 사용자나 아이템(콘텐츠)을 이용하여 추천한다. 마지막으로, 3세대는 사용자들의 명시적·암묵적 선호도 성향과 소셜 네트워크 정보를 이용하여 하이브리드 기법으로 동적으로 추천하는 방법으로 최근 이 소셜 정보를 어떻게 활용하여 적용할지 연구되고 있다. 본 논문은 이 3세대 추천 시스템을 기반으로 소셜 네트워크 정보 중 사용자 간의 친구 관계 같은 신뢰 정보를 추천 시스템에 적용한다.

일반적으로 추천 정보는 사용자들의 과거 선호도 즉, 사용자의 히스토리를 기반으로 가장 선호할 만한 Top-N개의 아이템(콘텐츠) 리스트를 제공한다 [3][4][5][6]. 이렇게 일방적으로 제공되는 추천 정보는 푸싱 (Pushing) 서비스로 사용자의 의견이 잘 반영되지 않는다. 그러므로 본 논문에서는 시각화 기술을 이용하여 동적으로 사용자들의 요구를 반영함으로써 사용자의 의사결정 능력을 향상시키고자 한다. 기존 추천 시각화의 경우 추천 리스트가 시각화 형태로 변경되는 것으로 이전에 제공되는 추천 방법과 다를 바가 없다 [7][8]. 즉, 시각화된 형태로 추천 리스트가 제공되는 것만 다르지 그 외에 사용자가 어떤 정보를 요구하는 등의 의사 결정 기회는 주어지지 않는다. 본 논문에서는 영화, 음악, 인물 등 다양한 토픽에 대한 정보를 구조화된 형태로 제공하는 거대한 협업적 지식 베이스인 Freebase와 연동을 통해 3가지 방향으로 시각화 정보를 제공한다. 첫째, 사용자가 추천 받고자 하는 N개의 아이템(콘텐츠) 수를 입력하면, 입력 받은 수만큼 해당 정보를 시각화 하여 제공한다. 예를 들어 영화를 추천하는 경우, 사용자가

추천 수를 '3' 입력하면 3개의 영화가 해당 포스터로 시각화하여 보여준다. 둘째, 앞에서 시각화된 아이템(콘텐츠) 중 특정 아이템(콘텐츠)을 선택하면 Freebase로부터 추가적인 정보를 시각화하여 제공한다. 영화인 경우 영화에 대한 부가정보 즉, 영화감독, 주연 배우, 장르 등이 제공되며, 그 외 인물이나 음악, 스포츠 뿐 만 아니라 위키피디아 등과 연결되어 풍부하게 부가 정보를 제공할 수 있다. 마지막으로 셋째 사용자들이 신뢰할 수 있는 사용자들 중 임의로 몇 명의 사용자를 선택하면, 이 사용자들의 성향을 기반으로 아이템(콘텐츠)을 추천해 준다. 이때, 사용자들은 클릭이나 드래그를 통해 사용자를 선택할 수 있으며 재 클릭 시 선택 해제가 된다.

본 논문의 구성은 다음과 같다. 2장에서는 신뢰 기반의 사용자 모델링을 기술하고, 3장에서는 본 논문에서 제안하는 추천 시각화를 기술한다. 그리고 4장에서는 추천 시각화를 위한 구현 및 시나리오를 기술하고, 마지막으로 5장에서 결론 및 향후 연구로 마친다.

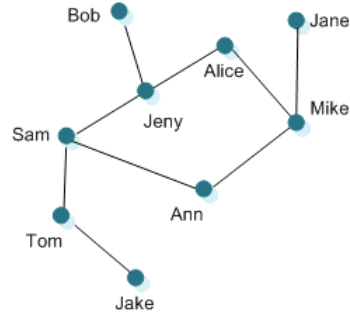
II. 관련 연구

2.1. 신뢰 기반의 사용자 모델링

먼저 사용자가 평가한 아이템 즉, 영화에 대하여 최소 1 ~ 최대 5 점으로 선호도 점수를 입력한 평가 정보를 이용하여 사용자-영화 매트릭스를 구성한다. 그리고 사용자가 명시적으로 신뢰를 표시한 링크 데이터를 기반으로 사용자-사용자 신뢰 네트워크를 구성한다. 구성된 사용자-영화 매트릭스에서 사용자가 평가한 영화를 기반으로 각 사용자들 간의 유사도를 수식(1)을 이용하여 계산한다.

$$sim(U_a, U_i) = \frac{\sum_{j \in \text{Commonly Rated Items}} (r_{aj} - \bar{r}_a)(r_{ij} - \bar{r}_i)}{\sqrt{\sum_{j \in \text{Commonly Rated Items}} (r_{aj} - \bar{r}_a)^2 \sum_{j \in \text{Commonly Rated Items}} (r_{ij} - \bar{r}_i)^2}} \quad (1)$$

수식(1)에서 U_a 는 목적 사용자, U_i 는 이웃하는 사용자 집단이고, r_a 는 목적 사용자가 평가한 아이템 점수, \bar{r}_a 는 목적 사용자가 평가한 아이템의 평균값이다. 그리고 사용자-사용자 신뢰 네트워크에서 각 사용자 사이의 간선 가중치 값을 이용하여 사용자간의 관계 정도를 계산한다. 관계 정도는 목적 사용자가 이웃 사용자들과 몇 명의 사용자들을 통해 연결되어 있는지를 말하며, 아래 [그림 1]와 같이 계산 할 수 있다.



(a) 사용자들의 신뢰 네트워크
(a) Trust-network of users

관계정도	Jake	Tom	Sam	Jeny	Bob	Alice	Ann	Mike	Jane
Jake	-	1	2	3	4	4	3	5	6
Tom	1	-	1	2	3	3	2	4	5
Sam	2	1	-	1	2	2	1	3	4
Jeny	3	2	1	-	1	1	2	2	3
Bob	4	3	2	1	-	2	3	3	4
Alice	4	3	2	1	2	-	2	1	2
Ann	3	2	1	2	3	2	-	1	2
Mike	5	4	3	2	3	1	1	-	1
Jane	6	5	4	3	4	2	2	1	-

(b) 신뢰 네트워크상의 관계 정도
(b) Social-relation of users

그림 1. 사용자들의 신뢰 네트워크와 관계 정도
Fig. 1. Trust-network and social-relation of users

사용자들의 신뢰 정보는 0과 1로 표기하며, 사용자 Jake가 사용자 Tom을 신뢰하는 경우 (Jake, Tom, 1)로 표현한다. 사용자들이 명시적으로 표시한 이 데이터를 가지고 [그림 1](a)의 사용자 신뢰 네트워크를 생성하고, 쉽게 확인할 수 있도록 (b)의 매트릭스로 표현한다.

일반적으로 얼마나 많은 사용자들과 관계를 맺느냐에 따라서 특정 사용자들 간의 관계를 계산하기 위한 경로는 여러 개 존재할 수 있다. 예를 들어, [그림 1]의 사용자 Sam과 Mike를 연결하기 위한 경로는 (Sam, Jeny, Alice, Mike)와 (Sam, Ann, Mike)로 2개가 존재한다. 이 여러 개의 경로들 중에서 어떤 경로를 선택하느냐에 따라 사용자간의 관계 가중치 값도 크게 달라진다. 그러므로 본 논문에서는 두 사용자 사이에 존재하는 모든 경로를 구하고, 경로에 존재하는 모든 사용자간의 표준편차를 계산한다. 이 표준편차 값을 이용하여 신뢰 관계 기반의 사용자 유사도를 아래 수식(2)으로 계산한다 [9].

$$SSim(u_i, u_j) = sim(u_i, u_j) + (1 - sim(u_i, u_j))(w_r) \quad (2)$$

수식(2)는 사용자 i와 j의 사용자 유사도 값을 구하는 것으로, W_{ij} 는 사용자-영화 매트릭스로부터 추출된 사용자간 유사도로 두 사용자가 관람한 영화의 평을 기반으로 한다. W_{ij} 은 두 사용자간의 관계 정도로 최종 선택된 사용자 경로에 따라서 아래 [표 1]의 관계 가중치 값을 의미한다.

표 1. 관계 정도에 따른 가중치
Table 1. Weight by relations

	간선 1	간선 2	간선 3	간선 4	간선 5	간선 6
가중치	1.0	0.8	0.6	0.4	0.2	0.0

2.2. Linked Data 기반의 영화 추천 시스템

Linked Data (LOD)는 시맨틱 웹의 창시자인 Tim Berners-Lee에 의해 정의된 것으로 "시맨틱 웹은 단지 데이터를 웹으로 제공하는 것이 아니라, 데이터 간의 링크를 만들어서 인간이나 기계 모두 데이터의 웹을 탐험할 수 있도록 해준다. Linked Data를 통해 유용한 데이터를 얻게 되면 그 데이터에 관계된 데이터로 계속 항해가 가능하다" 와 같이 설명하였다 [10].

Linked Data는 ① URI로 식별하여 연결(Linking)하고 ② RDF로 표현하며, ③ SPARQL로 질의하고 ④ HTTP로 유통한다는 4가지 기본 원칙을 지니고 있다 [11]. 이 Linked Data는 URI, RDF, HTTP를 통해 연결하여 사용할 수 있으므로 하나의 거대한 지식베이스처럼 사용할 수 있고 이를 통해 상호 운용성을 높일 뿐 아니라 데이터 통합을 용이하게 할 수 있다. 그리고 Linked Data를 통해 사용자가 원하는 데이터가 이미 존재하는지, 어디에 존재하는지를 알 수 있으므로 시스템의 사일로(silo)문제에 의해 발생된 불필요한 데이터 중복 문제를 해결할 수 있다. 뿐만 아니라 SPARQL Endpoint를 통해 SPARQL 질의가 가능하고 이를 응용 프로그램에서 이용할 수 있으므로 OpenAPI에 비해 데이터 접근을 구체화할 수 있어 데이터 지향의 매쉬업이 가능한 장점이 있다 [10]. 그러므로 이 Linked Data를 추천 시스템에 적용하여 보다 유용하고 다차원적인 추천 정보를 제공하고자 많은 연구자들이 연구 중에 있다 [12][13][14][15][16]. 이 연구들은 Linked Data를 영화 추천 시스템에 적용한 것으로, 보통 IMDB¹⁾, FreeBase²⁾, OMDb³⁾, DBpedia (Wikipedia) Movies⁴⁾,

RottenTomatoes.com, StanfordMovieDatabase⁵⁾ 등의 Linked Data를 이용하며 주로 LinkedMDB, FreeBase, DBpedia Movies를 사용한다. IMDB는 영화에 대한 최신 정보를 대량으로 제공하는 웹 상의 거대한 영화 데이터베이스로 RDF 타입으로 제공한다. FreeBase는 거대한 글로벌 지식 데이터베이스로 공유할 수 있으며 오픈 소스로 제공되어 쉽게 사용할 수 있다. 이 FreeBase는 "film" 카테고리에서 38,000개 이상의 영화 그리고 이 영화에 관련된 수천 개의 데이터로 구성되어 있다 [12]. DBpedia는 Wikipedia의 문서와 상응하는 36,000개의 영화에 대한 정보로 이뤄져 있고, OMDb는 9,000개의 영화에 대한 정보를 포함하는 영화 오픈 데이터 소스이다. 마지막으로 RottenTomatoes.com과 Stanford Movie Database도 공개된 영화 정보에 대한 데이터베이스이다. 이 Linked Data의 정보 이용은 일반적으로 영화 제목, 사람 이름(배우, 작가, 작곡가) 등 관련된 리소스는 owl:sameAs, rdfs:SeeAlso, foaf:page 등을 이용하여 추출이 가능하며, 사용자가 특정 온톨로지에서 정의한 predicate (user specific predicate)을 이용할 수 있다 [17]. 이 Linked Data를 적용한 추천 시스템을 [13][14][15]에서 구현하여 평가하였는데, [13]의 추천 시스템의 성능을 향상시키기 위해 협업적 여과 방법과 LOD 기반의 추천을 결합하여 제안한 경우 추천 시스템의 공통적인 이슈 사항인 콘텐츠 분석 한계, 초기 사용자 및 데이터 희소서의 문제를 완화시킬 수 있다. [14]에서는 소셜 네트워크 서비스 중 하나인 Facebook과의 결합을 통해 사용자 프로파일을 생성하여 Linked Data 기반의 추천 시스템에 적용하여 재현율과 정확도를 실험 평가함으로써 영화 추천에 유효함을 입증하였다. 마지막으로 [15]은 콘텐츠 기반의 추천 시스템을 위하여 모델 기반의 접근법을 사용한 연구로 아이템과 사용자 프로필 상의 양쪽 정보 표현을 LOD 데이터를 활용함으로써 앞서 언급한 콘텐츠 분석 한계 문제를 완화 시키는 것을 볼 수 있다. 즉, 현존하는 추천 시스템의 문제를 해결하기 위해 Linked Data를 적용하면 콘텐츠 정보가 풍부해져서 충분히 문제를 완화 시킬 수 있다.

2.3. 추천 시각화

사용자에게 추천 정보를 보다 효율적으로 제공하기 위해 추천 시스템의 시각화 연구가 지속적으로 진행되고 있다. 기존의 추천 방법은 콘텐츠 기반이나 협업적 여과 또는 하이브리드 방법의 사용자 모델링을 통하여 사용자가 평가하지 않은

1) IMDB: www.imdb.com
2) FreeBase: www.freebase.com
3) OMDb: www.omdb.org
4) DBpedia: www.DBpedia.org

5) Stanford Movie Database: infolab.stanford.edu/pub/movies

아이템이나 콘텐츠를 예측한 후, 예측된 값이 높은 N개의 아이템(콘텐츠) 리스트를 추출하여 해당 정보 (예를 들어, 영화 제목이나 책 제목 등)를 제공한다. 이때 제공되는 정보는 대표 이미지와 텍스트로 구성이 되며, 추천된 정보를 통해 사용자가 구매 또는 평가를 할 수 있다. 즉, 시스템이 제공해 주는 정보를 사용자는 확인 또는 평가만을 할 수 있다. 그래서 보다 사용자의 관심사를 높이고 사용자의 요구를 반영하기 위해 시각화 기술을 적용한 추천 시스템이 다양하게 연구되고 있다 [18][19][20][21][22][23]. Netflix⁶⁾, jinni⁷⁾, IMDB, MovieLens⁸⁾ 등 현존하는 영화 추천 시스템의 시각화 방법은 테이블(Tables)이나 선형 트리맵(Linear Treemap), 아이콘 타일 (Icon Tiles) 등의 형태를 주로 사용한다 [18]. 이 시각화 방법을 추천 시스템에 적용함으로써 사용자와의 상호작용(Interaction)을 통해 사용자들의 피드백을 받을 수 있고, 사용자에게 유용한 정보를 보다 효율적으로 제공할 수 있다. [18][19]은 사용자에게 추천 정보를 제공하기 위한 내비게이션(Navigation) 방법으로 연결 그래프 기반의 영화 추천 시각화를 제공한다. 이때 사용자가 영화를 평가한 점수(Rating), 영화 장르(Genre), 년도(Year) 등에 따라 여과(Filters)하여 정보를 제공받을 수 있다. [20]은 3개의 소스 - Wikipedia, Facebook, Twiteer - 에 대한 추천 방법으로 ① 가중치(weighted) ② 혼합(mixed) ③ 소스 교차(cross-source)의 하이브리드 메소드를 사용자들에게 시각화된 슬라이드 바를 이용하여 조절할 수 있다. 가중치 하이브리드 메소드는 각 소스에 대하여 사용자가 가중치 값을 설정하면, 각 소스를 위한 추천 점수의 가중치 합으로 나타낸다. 혼합 하이브리드 메소드는 각 소스의 추천 랭크 중 각각 top-n개씩 추출하고, 소스를 교대로 하나씩 추천해 준다. 마지막으로 소스 교차 하이브리드는 하나 이상의 소스에 나타나는 가장 선호하는 추천 접근법으로, 예를 들어 Facebook 소스는 협업적 여과로 Wikipedia는 콘텐츠 기반으로 추천을 생성해 내는 것이다. [21]는 시각화 기반의 협업적 여과 추천 시스템으로 시스템 내에 있는 데이터들과 시각적으로 인터랙션하고 보다 쉽게 이해할 수 있으며, 사용자들이 탐색(exploration)할 수 있도록 한다. 즉, 이웃하는 사용자들을 쉽게 판별할 수 있도록 하이라이트로 제공하고, 사용자가 아이템들을 선택하면 동적으로 선호도를 예측한다. 그리고 사용자를 선택하면 그 사용자가 평가한 점수도 공개하여 제공한다. 이 시각화 기반의 추천을 통해 협업적 여과가 가지고 있

는 초기 사용자 문제를 완화시키는데 도움이 될 수 있다. 그 외 사용자들의 행동 패턴을 기반으로 하는 시각화된 추천 방법[22]과 정보 시각화 기술을 통한 추천 시스템과의 인터랙션[23]을 통해 그래픽적인 표현 방법을 제공함으로써 추천 결과에 대한 사용자 신뢰도를 향상시키고, 사용자가 관심 갖을 수 있는 아이템을 찾는 능력을 증가시킨다.

전반적으로 추천 시스템에 시각화 기술을 적용하여 사용자에게 제공하는 경우, 사용자 인터페이스를 통해 사용자 만족도를 증가시킨다. 그리고 추천 할 때의 인터랙션을 통해 추천 정확도와 사용자 경험(user experience)도 같이 향상시킬 수 있다. 그러므로 본 논문에서는 추천 시스템 상에서 사용자의 의사 결정 능력을 향상시키기 위해 이 시각화 기술을 적용하고자 한다.

III. Freebase 기반의 추천 시스템 시각화

본 논문은 2.1의 사용자의 소셜 관계를 기반으로 생성된 사용자 모델링을 통하여 영화를 관련한 사용자에게 시각화된 추천 시스템을 제공하는 것이다. 이를 위해 사용자가 영화에 대하여 평가한 평가 정보와 사용자의 social relationship을 가지고 있는 이웃 사용자들의 영화 평가 정보를 바탕으로 평가되지 않은 영화를 추천해 준다. [그림 2]는 본 논문에서 제안하는 Freebase 기반의 추천 시각화 구조도이다. [그림 2]의 시스템 구조도와 같이 사용자의 요구에 따라 영화에 대한 다른 정보 탐색 (exploration) 뿐 아니라 이웃 사용자들을 선택하여 여기에 맞는 영화를 추천해 준다.

본 논문에서는 영화 추천 시각화를 위해 Linked Data 중 Freebase를 적용하여 영화에 대한 부가 정보를 제공한다. [그림 2]를 보면 먼저 사용자들이 평가한 영화를 기반으로 사용자 모델링을 하고, 성향이 유사한 사용자들을 하나의 군집으로 형성하여 이들을 기반으로 평가하지 않은 영화에 대해서 선호도를 예측한다. 선호도 예측 값은 내림차순으로 정렬하고, 시각화 인터페이스를 통해 입력받은 추천 수 N에 따라 상위 N개의 영화 리스트를 추출한다. 추출된 영화 리스트는 내부적으로 Freebase에 영화 제목으로 질의(query)를 보내고, 영화에 관련된 부가 정보 - 주연 배우, 영화 음악, 감독, 장르 등 - 를 가져온다. 추천 시각화는 총 3가지 방법으로 제공되며, ① 사용자가 제공받고자 하는 영화의 수만큼 영화 포스터로 시각화(Top-N Visualization)하고, ② 추천된 영화 중 특정 영화를 선택하면 영화감독, 주연 배우, 장르 등의 부가

6) Netflix: Netflix.com

7) Jinni: jinni.com

8) MovieLens: movielens.org

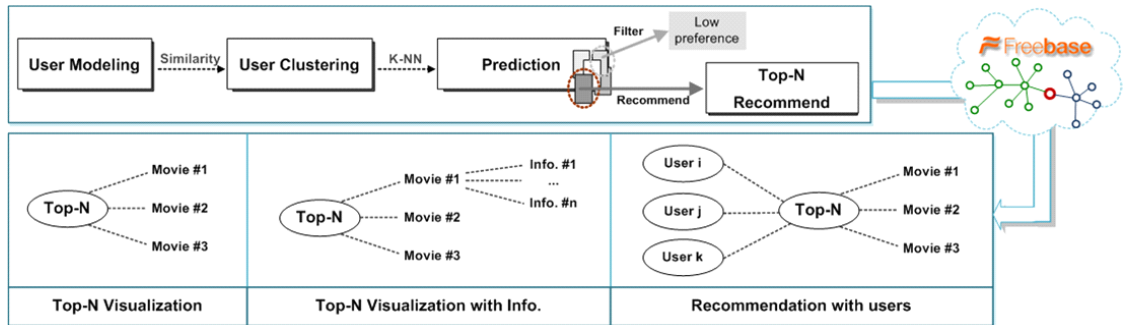


그림 2. 개인화된 추천 시각화 구조도
Fig. 2. Personalized recommendation visualization structure

적인 정보를 시각화하여 제공한다 (Top-N Visualization and Information). 마지막으로 ③ 신뢰 기반의 사용자들 중 임의로 몇 명을 이웃 사용자로 선택하여 추천한다 (Recommendation with users).

3.1. 아이템 추천 단계

2.1의 사용자 모델링을 기반으로 사용자가 평가하지 않은 영화에 대하여 선호도를 예측하는 방법은 다음과 같다. 먼저 사용자 모델링을 통해 사용자가 평가한 영화를 기반으로 사용자들 간의 유사도 값을 측정한다. 유사도 값은 협업적 여과 기반의 추천 시스템에서 가장 많이 사용되는 피어슨 상관 계수(Pearson Correlation Coefficient)를 사용하며 수식(1)과 같으며, 사용자간 유사도는 [-1,1]로 표현이 된다. 사용자 간의 선호 값이 많이 유사할수록 1에 가까우며, -1로 갈수록 사용자의 성향이 정반대임을 알 수 있다. 측정된 유사도 값을 기준으로 유사도 값이 높은 순으로 정렬하여 K명의 이웃 사용자 집단을 생성한다. 마지막으로 목적 사용자가 평가하지 않은 영화는 K명의 이웃 사용자들이 평가한 영화 선호도 값을 이용하여 수식(3)과 같이 예측한다.

$$P_{aj} = \bar{r}_a + \frac{\sum_i SSim(U_a, U_i)(r_{ij} - \bar{r}_i)}{\sum_i SSim(U_a, U_i)} \quad (3)$$

수식(2)에서 목적 사용자 a가 평가하지 않은 특정 영화 j에 대하여 선호도를 예측하기 위해, 공통적으로 평가한 영화에 대하여 K명의 이웃하는 사용자 i의 평가 값과 그 사용자와의 유사도 값 $sim(U_a, U_i)$ 을 이용하여 예측한다. 예측된 평가 값은 내림차순으로 정렬하여 추천 목록을 생성하고, 사용자가 시각화 인터페이스를 통해 추천 수 N을 입력하면 그

에 맞는 상위 N개의 추천 목록을 제공한다 [24].

3.2. Freebase 기반의 추천 시각화

사용자의 선호도에 맞는 추천 목록이 추출되면 사용자의 요청에 따라 영화 리스트가 시각화하여 제공되기도 하고, 영화에 대한 부가 정보까지 제공할 수 있다. 영화의 부가 정보는 거대한 협업적 지식 베이스인 Freebase를 이용하여 제공한다. Freebase는 주제(subject) 또는 엔티티(entity)에 대한 정보 및 개념(concept)으로 구성되어 있으며, 각 엔티티는 유일한 ID로 구분한다. 그리고 주제는 하나 또는 그 이상의 타입(type) 및 속성(property)으로 되어 있어서 SPARQL과 같은 MQL (Metaweb Query Language)을 이용하여 질의를 할 수 있다. MQL은 트리 기반의 결과와 함께 객체 지향적인 질의 언어로 쉽게 사용이 가능하다[25]. 그러므로 URI 검색과 효율적인 브라우징 제공을 위해 Freebase API를 매쉬업하고, 표준 HTTP 요청과 응답을 통해 JSON (JavaScript Object Notation)기반의 질의를 한다. 영화 정보를 위해 Freebase의 Film 도메인을 이용하며, 이 Film 도메인은 54개의 타입과 1,989,201개의 인스턴스들로 구성되어 있다. 그리고 33,000개 이상의 영화와 위키피

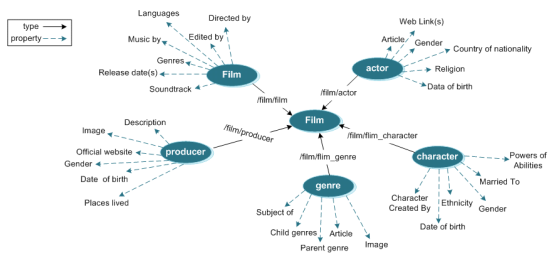


그림 3. Freebase의 Film 도메인에 대한 타입(Type)과 속성(Property)
Fig. 3. Type and Property about Film domain of Freebase

디아로부터 로딩된 23,000개의 구조화된 정보들로 구성되어 수 천명의 연기자, 디렉터, 프로듀서 등 사용자에게 보다 풍부한 정보 제공이 가능하다. Freebase의 Film 도메인에서의 타입과 속성들은 아래 [그림 3]과 같으며, 이 외 50개의 타입과 각 타입에 대한 속성들이 존재하므로 [그림 3]에 있는 타입과 속성들로 한정하지 않는다.

Freebase 기반의 영화 추천 시스템을 시각화 기술과 접목하여 본 추천 시스템을 이용하는 사용자의 의사 결정 능력을 향상시키기 위해서 아래 3가지 방법으로 시각화하여 제공한다. 먼저 ① 사용자 유사도를 통해 추출된 Top-N개의 영화 리스트를 시각화하여 제공한다. 이때 N은 사용자의 요구에 따라 동적으로 변동이 가능하며, 사용자의 요청에 맞게 추천된 영화 항목들을 나타낸다. ② 영화에 대한 정보 속성을 이용하여 사용자에게 부가적인 정보를 시각화하여 제공한다. [그림 3]과 같이 Endpoint로 연결된 Freebase로부터 영화에 대한 시맨틱 정보를 포함하기 때문에, 앞서 추천된 영화를 사용자가 선택하면 Freebase로부터 부가 정보를 추출하여 보여준다. 마지막으로 ③ 시스템과 인터랙션하는 사용자가 이웃 사용자를 선택하면, 이들을 기반으로 적합한 영화 항목을 추천한다. 각 시각화를 위한 자세한 설명은 다음 절에서 기술한다.

3.2.1. Top-N 추천 시각화

2.1절의 사용자 모델링과 3.1절의 영화 추천 단계를 통하여 영화에 대한 선호도 예측 값이 추출되면, 사용자는 몇 개의 영화를 추천 받을지 시스템에 요청할 수 있는 폼(입력 창)이 [그림 4]와 같이 생성된다.

Top-N 추천 시각화는 사용자로부터 직접 추천 받을 수를 입력 받고 상위 아이템 즉, 사용자의 선호도 예측 값이 높은 영화부터 차례대로 나타내준다. 영화의 예측 선호도는 수식 (3)을 이용하여 계산하며, 여기서 이웃 사용자 집단은 사용자 간 유사도 값이 높은 사용자들을 이용한다.

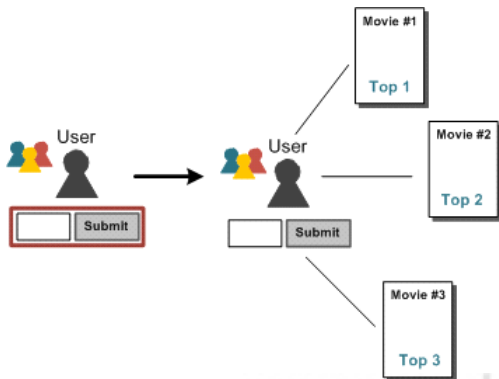


그림 4. Top-N 추천 시각화
Fig. 4. Top-N Recommend visualization

시각화하여 추천 목록을 제공하는 것은 영화 이미지 또는 영화 포스터를 시각화하여 사용자에게 제공하며, 영화에 관련된 이미지와 포스터는 Freebase와의 연동을 통해 /film/producer/image 또는 /film/film_genre/image로부터 추출이 가능하다.

표 2. Freebase의 Film에 대한 기본 속성
Table 2. Basic attributes of film on Freebase

Genre	Actor	Character
Runtime	Soundtrack	Director
Producer	Writer	Editor

3.2.2. 정보 속성 기반의 Top-N 추천 시각화

3.2.1의 Top-N 추천 시각화에서 사용자가 특정 영화를 선택하면, Freebase로부터 해당 영화에 대한 속성을 추출하며, 본 논문에서는 [표 2]와 같은 기본 속성을 주로 이용한다.

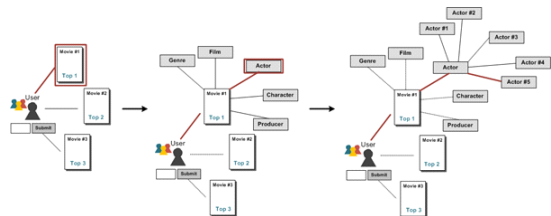


그림 5. 정보 속성 기반의 Top-N 추천 시각화
Fig. 5. Top-N recommendation visualization based on Information attributed

기본 속성을 이용한 시각화 방법은 아래 [그림 5]와 같다. 먼저, Top-N 추천 시각화와 똑같이 사용자로부터 추천 받을 영화의 수를 입력하면 입력 받은 수만큼 영화 목록을 제공한다. 이때 사용자가 특정 영화를 하나 선택을 하면, 그 영화가 가지고 있는 기본 속성 - Genre, Film, Actor, Character, Producer - 등이 메뉴처럼 시각화된다. 이 중 한 속성을 선택하면 해당하는 인스턴스들이 출력되고, 사용자가 한 인스턴스를 선택하면 그와 관련된 자세한 정보를 새 창을 통해 제공해 준다.

예를 들어, A 사용자가 3개의 영화를 추천 받고자 '3'을 입력하여 전송한 후, 3개의 영화 '더 울버린, 레드: 더 레전드, 티보'를 추천 받는다. 추천 받은 영화 중 '더 울버린'을 선택하면, 기본 속성인 '배우, 감독, 장르, 상영시간' 등이 제공된다. 사용자가 '배우'를 선택하면 '더 울버린'에 나오는 배우들 '휴 잭맨, 제시카 비엘, 윌 윤 리, 사나다 히로유키, 브라이언 티 등 배우 리스트가 나온다. [그림 5]는 여기까지의 단계를 나

다낸 것이고, 만약 사용자가 배우 '휴 잭맨'을 선택하면 휴 잭맨에 대한 기본적인 내용 - 생년월일 등 개인 기본 정보, 출연 영화 목록(관련 비디오) 등 - 들이 새 창(또는 팝업)을 통해서 제공된다. 이 Linked Data는 각 타입과 속성들이 서로 연관되어 관계를 지니고 있기 때문에 사용자의 요청에 따라 지속적으로 탐색(exploratory search)가 가능하다.

3.2.3. 사용자 기반의 Top-N 추천 시각화

앞의 2가지 추천 시각화 방법은 현재 활성화된 사용자를 기준으로 추천 목록을 추출하여 제공하지만, 본 절에서는 선호도 예측을 할 때 필요한 이웃 사용자들을 사용자가 직접 선택할 수 있다. 이웃 사용자 선택은 [그림 6]과 같이 사용자가 추천 수를 입력하는 곳에 있는 3명의 사용자가 같이 있는 아이콘을 이용한다. 이 사용자 아이콘을 선택하면 관련 있는 여러 사용자들이 출력되는데, 시스템 내에서 활성화된 사용자가 신뢰 관계를 표시한 사용자들로 구성이 된다. 이때 시스템에 있는 모든 사용자들이 아니라, 2.1절의 신뢰 기반의 관계 정도를 이용하는 것으로 본 논문에서는 관계 정도 3까지만 고려한다. 관계 정도가 4 이상이 되면 두 사용자 사이를 연결하기 위해 최소 4명 이상의 다른 사용자를 거쳐야 되는 것으로 이 경우 두 사용자의 성향이 상반되는 것을 알 수 있다.

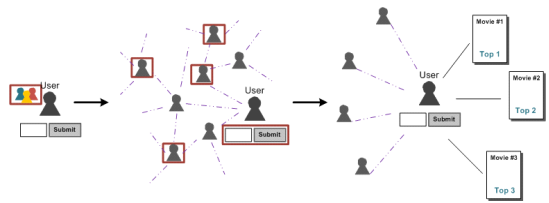


그림 6. 사용자 기반의 Top-N 추천 시각화
Fig. 6. Top-N Recommend visualization based on a user

[그림 6]과 같이 활성화된 사용자가 관계 정도 3이내에 있는 사용자들 중 추천을 위한 이웃 사용자들로 선택을 하면, 활성화된 사용자와 선택된 이웃 사용자간의 사용자 유사도와 영화 평가 값을 이용하여 추천 목록을 동적으로 생성한다. 그리고 활성화된 사용자가 추천 받을 영화 수를 입력하면 입력 값과 일치하는 수의 영화 목록을 3.2.1절과 3.2.2절 같이 시각화하여 제공해 준다.

IV. 구현 및 시나리오

본 논문에서 제안하는 Freebase 기반의 추천 시스템 시각화를 위하여 Intel Core i7-2600 3.40GHz, RAM 6GB,

MS Windows 7 MS-SQL 2005 환경에서 ASP.NET with C#, HTML, CSS, JavaScript로 구현하였다. 먼저 영화에 대한 정보를 수집하고 활용하기 위하여 아래 [그림 7]과 같이 Freebase에서 지원하는 MQL을 이용한다.

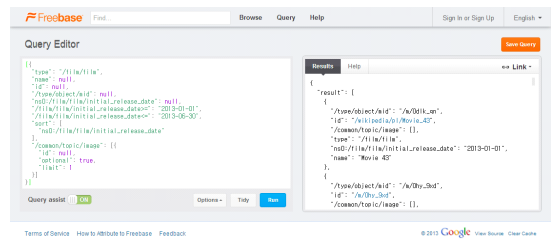


그림 7. Freebase 기반의 MQL Query Editor
Fig. 7. MQL Query Editor based on Freebase

예를 들어, [그림 7]의 왼쪽 Query Editor에 Freebase의 토픽인 "/flim" 중 등록된 최초 상영일 (Initial_release_date)를 기준으로 2013년 1월 1일부터 2013년 6월 30일까지의 영화를 질의(query)하면, 오른쪽 결과 화면에 영화 1006편에 대한 기본 정보 즉, 영화 ID, 영화 이미지, 영화 제목 등을 가져온다. 가져온 이 정보는 사용자에게 무조건 제공되는 것이 아니라, 시각화된 사용자 인터페이스를 통해 요청된 정보만 제공된다. 이를 위해 영화에 대한 이 기본 정보를 JSON 형식으로 받아 [그림 8]의 tbMovie 테이블에 저장하고, 내부적으로 Freebase에서 정보를 가져와 테이블에 저장하기 하여 .NET C#으로 웹크롤러(Webcrawler)를 구현하였다.

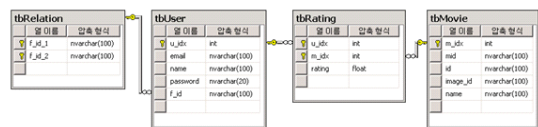
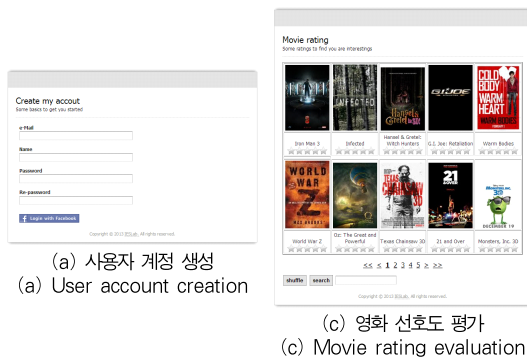


그림 8. 데이터베이스 스키마 다이어그램
Fig. 8. Database schema diagram

그리고 사용자의 성향에 맞는 영화를 추천하기 위하여 [그림 9]와 같이 웹 사이트를 구현하였고, 이를 통해 사용자가 명시적으로 영화 선호도를 평가할 수 있도록 하였다. 먼저 [그림 9](a)에서 보는 것과 같이 사용자의 기본 정보 - 이메일, 이름, 비밀번호 - 를 입력 받아 사용자 계정을 생성하고, 생성된 개인 정보는 [그림 8]의 tbUser 테이블에 저장된다. 이메일은 Facebook 계정을 입력하며, Facebook 계정으로부터 사용자의 기본 정보 - 사용자 메인 이미지, 친구 목록 - 를 가져

온다. 이때 [그림 9](a)의 Facebook 계정으로 로그인을 선택하면 보다 빠르고 편리하게 입력할 수 있으며, Facebook 계정과의 연동을 통해 [그림 9](b)의 Graph API Explorer로 친구 목록을 추출할 수 있다. 사용자 계정을 통해 로그인하면 영화에 대한 선호도를 평가할 수 있는 페이지가 [그림 9](c)와 같이 제공된다. 구체적으로 영화 선호도 평가를 위한 페이지는 Freebase에 등록된 영화 목록을 보여주고, 선호도 점수를 1 ~ 5 사이의 값으로 선택할 수 있도록 한다. 선호도는 1에 가까울수록 사용자가 영화를 비 선호 즉, 싫어하는 영화라 볼 수 있고 반대로 5에 가까울수록 매우 선호하는 것으로 판단한다. 사용자들이 명시적으로 평가한 영화에 대한 선호도 점수는 [그림 8]의 tbRating 테이블에 사용자ID와 영화ID 그리고 평가 값으로 저장된다. 그리고 추천 시각화는 정적인 웹 페이지에서 동적으로 사용자의 요구를 반영하기 위해 AJAX (Asynchronous JavaScript and XML) 기술을 적용하였으며, 이를 통해 추천된 영화 정보와 영화 별 상세 정보를 실시간으로 가져올 수 있도록 구현하였다.



(a) 사용자 계정 생성
(a) User account creation

(c) 영화 선호도 평가
(c) Movie rating evaluation

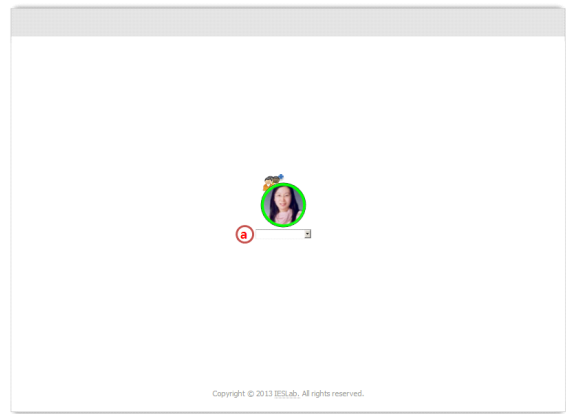


(b) Graph API Explorer를 통한 친구 목록 추출
(b) Collect friend list via Graph API Explorer

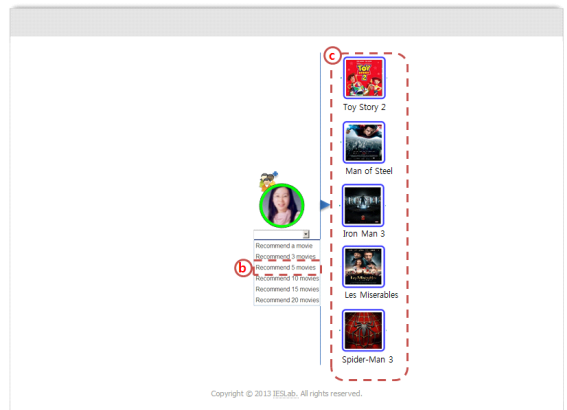
그림 9. 사용자의 영화 선호도를 수집하기 위한 웹 페이지
Fig. 9. Webpages for collect to movies ratings of a user

본 논문에서 제안하는 시각화 구현을 위한 시나리오는 3.2.1절 ~ 3.2.3절에 기술한 내용을 바탕으로 3개의 관점에서 기술한다. 첫째, Top-N 추천 시각화로 [그림 10]에서와 같이 로그인한 사용자를 중심으로 영화 목록을 추천 받을 것

인지, 아니면 신뢰할 수 있는 이웃 사용자 (연결 정도 3 이내 인 친구) 목록을 제공받을 지 선택할 수 있다. 영화 목록을 추천 받기 위해 화면에서 제공되는 콤보 박스를 선택하면 사용자가 제공받을 추천 수를 선택할 수 있고, 기본적으로 첫 화면에 Top-1부터 Top-10까지의 항목을 제공한다. 그 이상의 영화를 추천 받고자 하는 경우 콤보 박스의 슬라이드를 조절하면 Top-N개를 선택할 수 있다.



(a) 로그인 후 첫 화면
(a) Main screen after user login



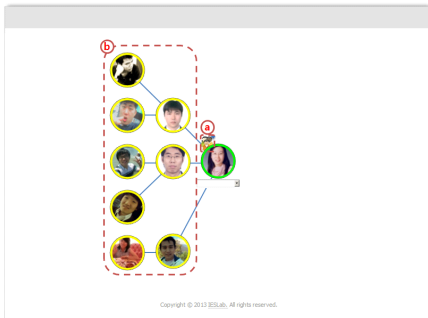
(b) Top-N 선택 시 해당 추천 목록 제공
(b) Top-N recommended Movie list

그림 10. Top-N 추천 시각화
Fig. 10. Top-N recommendation visualization

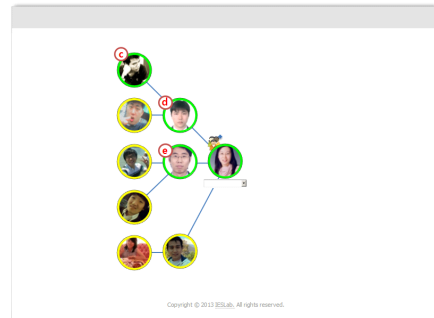
예를 들어, [그림 10]과 같이 사용자 Inay가 제안하는 시스템에 로그인을 하면 등록된 사용자의 이미지와 함께 추천 수를 선택할 수 있는 콤보 박스가 제공된다. 콤보 박스에서 5개 추천 항목을 선택하면, 선호도 예측 값이 높은 영화 항목

표 3. Freebase의 Film 도메인의 타입(Types)과 속성(Properties)
Table 3. Types and properties of Film domain of Freebase

Type	Properties				
Common/Topic	Also known as	Official website	Topical webpage	Web Link(s)	Notable for
	Description	Topic equivalent webpage	Subjects	Properties	Notable types
	Image	Social media presence	Subject of	Weblink	
Film	Initial release date	Story by	Runtime	Featured Locations	Metacritic film ID
	Tagline	Cinematography	Languages	Film Series	Apple movie trailer ID
	Directed by	Edited by	Country of origin	Sequel	Rotten Tomatoes ID
	Performances	Release date(s)	Notable filming locations	Prequel	Trailer Addict ID
	Personal appearances	Casting director	Gross revenue	Film format	Fandango ID
	Dubbing performances	Production design by	Estimated budget	Production companies	Pre-production
	Genres	Art direction by	Rated	Distributors	Filming
	Subjects	Set Decoration by	Trailers	Other film companies	Post-production
	Executive Produced by	Costume design by	Film Collections	Netflix ID	Songs
	Produced by	Other crew	Soundtrack	Film festivals	
	Screenplay by	Music by	Featured Song	NY Times ID	
Awards	Award Nominations		Awards Won		



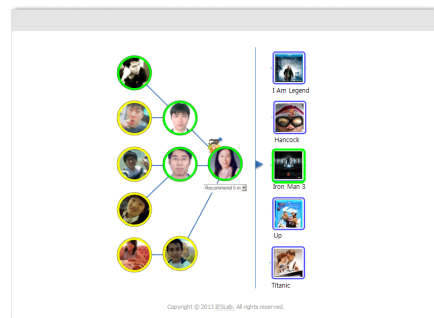
(a) 친구 목록 확장
(a) Extend friend list



(b) 이웃 사용자로 선정
(b) Select neighborhood users



(c) 선정된 사용자 기반으로 Top-N 선택
(c) Select Top-N by a chosen user



(d) 선택한 N개의 추천 목록
(d) Top-N recommended list

그림 12. 사용자 기반의 Top-N 추천 시각화
Fig. 12. Top-N recommendation visualization based on a user

순으로 위에서 아래 방향으로 영화 이미지와 제목을 시각화하여 제공한다. 사용자 Inay는 가장 선호 예측 값이 높은 영화가 Toy Story 2가 되며, Spider-Man 3는 5번째로 추천된 항목임을 알 수 있다. 사용자의 이미지는 Facebook API를 통해 추출하여 사용하며, 영화 이미지는 Freebase의 영화 이미지 속성을 추출하여 제공한다.

일반적으로 사용자에게 유용한 추천 항목은 10개 이내로 그 이상이 되면 제공되는 정보의 수가 많아져 오히려 불편한 정보 제공으로 변질될 수 있다. 그러므로 본 논문에서는 기본적으로 제공되는 Top-N개의 수를 10개 이내로 제공하며, 그 이상을 원하는 사용자를 위해 슬라이드 막대로 그 이상의 수를 선택할 수 있도록 하였다. 또한 정보를 제공받기 위해 화면을 보다가 키보드로부터 입력 값을 받으면 사용자에게 불편함을 제공할 수 있어 키보드로부터의 입력 값을 최소화하고 사용자가 마우스를 통해 선택할 수 있도록 시각화 UI를 설계하였다.

둘째, 앞에서 추천된 영화 목록에서 확장된 시각화 방법으로 (그림 11)과 같이 추천 영화 중 특정 영화를 선택하면 그 영화가 가지는 기본 정보인 감독, 제작사, 배우, 작가, 장르, 상영시간 등 영화의 속성 값이 제공된다.

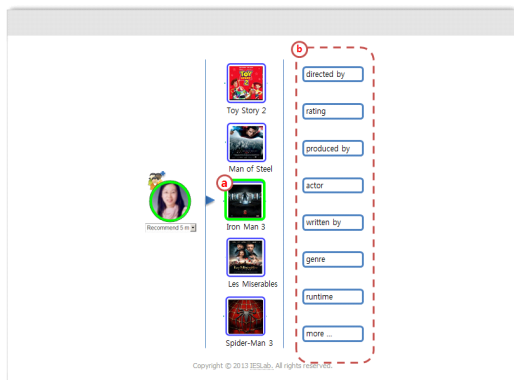
예를 들어, 앞의 사용자 inay가 추천된 항목 중 Iron Man 3 영화를 선택하면 이 영화의 속성 값이 시각화되어 사용자가 특정 정보를 요청할 수 있도록 제공한다. 이때 부가정보 중 배우(actor)를 선택하면 Robert Downey JR. / Guy Pearce / Cobie Smulders / Gwyneth Paltrow 등 배우 목록을 제공하고, 특정 배우를 선택하면 그 배우에 관련된 자세한 정보가 새 창(팝업창)에 제공된다. 이 영화에 대한 부가 정보는 (그림 7)의 Freebase MQL을 통하여 JSON 형태로

영화에 대한 속성 정보를 추출하며, 속성에 해당하는 인스턴스에 대한 상세 정보는 Freebase와 연결된 Wikipedia를 통해 해당 페이지로 연결되어 제공된다. Wikipedia와의 연결은 Freebase에서 타입(Types) 중 Common 타입으로 Topic equivalent webpage 속성으로 연결된다. 사용자 inay가 선택한 Iron Man 3에 대한 Freebase 정보는 Common, Film, 그리고 Awards로 크게 3가지 타입으로 구성된다. Common 타입 중 Topic(/common/topic)은 아래 [표 3]과 같이 14개의 속성 값으로 구성되며, Film은 53개의 속성 그리고 Awards는 2개의 속성으로 구성되어 있다.

[표 3]에서 영화 배우에 대한 속성은 Film 타입의 Performances 속성이며, /film/film/starring에서 Actor, Character, Special Performance Type 그리고 Character note에 대한 인스턴스들을 포함한다. Iron Man 3의 경우 아래 [표 4]와 그 외 12명의 배우에 대한 인스턴스들을 가지고 있다.

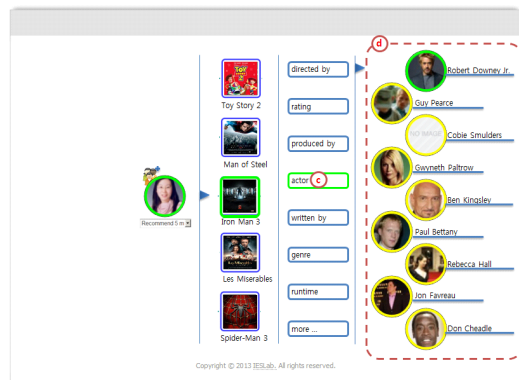
표 4. Iron Man3의 Performances 속성
Table 4. Performances attributes of Iron Man3

Actor	Character	Special Performance Type	Character note
Robert Downey Jr.	Iron Man	-	-
Guy Pearce	Aldrich Killian	-	-
Gwyneth Paltrow	Pepper Potts	-	-
Ben Kingsley	The mandarin	-	-
Paul Bettany	Jarvis	Voice	-
Rebecca Hall	Maya Hansen	-	-
Jon Favreau	Happy Hogan	-	-



(a) 영화 선택 시 기본 속성 제공

(a) Movie attributes list to a chosen a movie



(b) 특정 속성 선택 시 해당 인스턴스 제공

(b) Instances according to a chosen attribute

그림 11. 정보 속성 기반의 Top-N 추천 시각화
Fig. 11. Top-N recommendation visualization based on Information attributed

마지막으로 사용자 기반의 Top-N 추천 시각화를 위해 등록된 사용자의 이메일을 통해 Facebook 상에 친구로 등록된 친구 목록을 추출한 후, 시각화하여 제공한다. 일반적으로 Facebook은 직접적으로 연결된 친구들의 목록만 가지고 있을 수 있으나, 본 논문에서 사용한 2.1절의 신뢰 기반 사용자 모델링을 통해 관계 정도 3단계 사용자 즉, 친구의(1단계) 친구의(2단계) 친구(3단계)까지 고려하여 제공한다. (그림 12)와 같이 사용자가 친구 아이콘을 선택하면 3단계 이내의 사용자 목록이 제공된다. 제공된 사용자 목록 중 특정 사용자 (b)와 같이 선택을 한 후, (c)처럼 Top-N을 선택하면, 선택된 사용자들의 영화 선호도 성향을 기반으로 N개의 영화 목록을 제공한다.

시각화 방법 중 첫 번째, 두 번째는 사용자가 평정한 영화 정보를 기반으로 각 사용자들과의 영화 선호 성향(유사도)을 계산하여, 선호하는 성향이 유사한 사용자들을 이웃 사용자로 선정한다. 그리고 선정된 이웃 사용자를 기반으로 N개의 영화 추천 목록을 생성하여 사용자에게 제공하는 방식이다. 하지만, 이 세 번째 시나리오는 사용자가 임의로 이웃 사용자를 선택하여 이 사용자들의 영화 선호 성향만을 고려한 추천 목록을 생성하게 된다. 예를 들어, 사용자 inay가 주말에 3명의 친구와 집에서 영화를 다운받아서 보기로 한 경우, 시스템에 접속하여 함께 영화 보기로 한 친구 3명을 친구 목록에서 선택한다. 즉, 사용자 inay와 영화 보기로 한 친구 John과 Bob 그리고 Bob의 친구인 Nurzid를 친구 목록에서 선택하고, Top-5의 영화 추천을 요청한다. 그러면 시스템은 inay, John, Bob 그리고 Nurid의 영화 선호도 성향을 바탕으로 영화에 대한 선호도를 예측한다. 예측된 값을 토대로 선호도 값이 높은 영화 순으로 정렬하여 상위 5개의 영화를 추천 목록으로 생성하여 제공한다. 이후 영화 선택 및 영화에 대한 부가 정보 요청과 같은 인터랙션은 앞의 첫 번째, 두 번째 시나리오 방법과 동일하다.

사용자의 프로필을 수집하고 그 수집된 정보를 바탕으로 추천한 결과를 일방적으로 푸싱해주는 서비스로 사용자의 현재 의견이 잘 반영되지 않는 기존 추천 시스템과는 다르게 본 논문에서는 일방적으로 푸싱하지 않고 시각화 기술을 이용하여 동적으로 사용자들의 현재 요구를 반영함으로써 사용자의 의사결정에 따른 현시점의 요구사항을 반영한 추천결과를 제공할 수 있었다.

V. 결론 및 향후 연구

일반적으로 제공되는 추천 시스템은 사용자들의 히스토리

를 기반으로 가장 선호할 만한 Top-N개의 아이템(콘텐츠) 리스트를 제공해 주는 방식이다. 이와 같은 추천은 일방적으로 제공되는 추천 정보로 푸싱 (Pushing) 서비스와 같이 사용자의 의견이 잘 반영되지 않는다. 그러므로 본 논문에서는 사용자들의 요구를 반영하는 추천 시스템을 위해 시각화 기술을 적용한 영화 추천 시스템을 제안하여 동적으로 사용자들의 요구를 반영하고 사용자들의 의사 결정 능력을 향상시키고자 한다. 사용자들에게 영화에 대한 부가 정보를 제공하기 위해 구조화된 형태로 정보를 제공하는 Freebase와 연동하여 3가지 방법으로 시각화 정보를 제공한다. 첫째, 사용자가 추천 받고자 하는 N개의 영화 수를 입력하면, 입력 받은 수만큼 영화 목록을 시각화하여 제공한다. 둘째, 앞에서 시각화된 영화 목록 중 특정 영화를 선택하면 Freebase로부터 부가적인 정보 - 영화 감독, 주연, 배우, 장르 등 - 를 시각화하여 제공한다. 마지막으로 사용자들이 신뢰할 수 있는 사용자들 중 임의로 몇 명의 사용자를 선택하여 이웃 사용자로 설정하면, 이 사용자들의 성향을 기반으로 영화를 추천해 준다. 즉, 본 논문에서 제안하는 Freebase 기반의 영화 추천 시스템은 추천 수나 이웃 사용자 선택을 통해 동적으로 사용자들의 요구를 반영할 수 있고, Linked Data와의 연동을 통해 보다 풍부하게 부가 정보를 제공받을 수 있다.

본 논문에서 제안하는 추천 시각화 방법은 기존 시스템들 - Facebook, Freebase, Wikipedia 등 - 과의 연동을 통하여 웹에 존재하는 다양한 정보들을 재사용할 수 있다. 그리고 대량의 정보들 중 사용자가 선호할 수 있는 정보들을 선별하여 제공함으로써, 사용자가 보다 양질의 정보를 접할 수 있는 기회가 주어진다. 뿐만 아니라 추천된 정보에 대하여 시각 기법을 통해 상세한 정보를 추출하여 제공함으로써 사용자의 의사 결정 능력을 향상시킬 수 있으며, 보다 유용한 정보로 사용자를 유도할 수 있다. 시각화 정보는 영화 추천인지 이웃 사용자 선택인지 직관적으로 쉽게 판단할 수 있도록 간단하게 아이콘과 콤보 박스를 제공함으로써 선택에 대한 고민을 최소화 하였다.

본 논문은 영화 도메인을 선정하여 구현하였으나 제안하는 추천 시각화는 영화라는 도메인에 한정된 것은 아니고, 여행이나 음식점(레스토랑) 또는 도서 등 다양한 도메인에 적용할 수 있다. 예를 들어, 친구들과 여행을 가기 위해 여행지를 선택할 때 친구들의 성향을 고려하여 최대한 만족할 수 있는 여행지를 선택하는 것이 어렵다. 또한 지인이나 가족들과 음식점을 선택할 때도 모두를 만족시킬 수 있는 음식점 선택이 매우 어렵다. 이때, 본 논문에서 제안하는 시각화 추천 기법을 통해 보다 효율적인 선택을 할 수 있도록 도움을 줄 것을 기

대한다. 뿐만 아니라 스마트 시대 또는 빅데이터 시대로 접어들면서 블로그, 트위터 등 소셜 네트워크 플랫폼이 활성화되면서 다양한 정보 및 콘텐츠가 폭발적으로 증가하고 있다. 이 스마트 데이터 또는 빅데이터를 본 논문에서 제안하는 추천 시각화에 적용함으로써 고립될 수 있는 사용자의 유용한 정보를 제공할 기회가 확대될 수 있다. 그러므로 향후 연구로 추천 시각화를 위한 사용자 데이터를 이 빅데이터를 이용하여 다양한 도메인 - 영화, 여행, 음식점(레스토랑), 도서 등에 적용해 봄으로써 시각화 기법을 통한 사용자 의사 결정 능력이 향상되는지 실험해 볼 예정이다. 그리고 빅데이터와 같은 사용자 데이터는 시간의 흐름에 따라 급변하기 때문에 순차 패턴과 같이 시간에 대한 정보를 분석하여 동적으로 사용자의 선호 성향을 반영하는 사용자 모델링을 통한 추천 시각화 방법을 향후 연구할 계획이다. 또한 기존 추천 시스템의 추천 결과와 본 논문에서 제안하는 시각화를 통해 사용자가 원하는 이웃사용자를 직접 선택하고 그것을 기준으로 얻어지는 추천 결과를 정량적인 비교가 필요하다. 끝으로 본 논문에서는 추천 시각화를 위해 3가지 기법을 이용하였는데, 추천된 영화를 기준으로 선호도가 높을 것으로 예상되는 사용자들을 추천하거나 사용자가 평소 선호하는 영화를 장르 또는 배우와 같은 속성별로 수치화하여 구글 차트(Google Charts)와 같은 기법을 통한 시각화 방법 연구도 필요하다.

참고문헌

- [1] Se-yeoung Park, Hyeon-Jae Lee, Jun-Myeon Jo, "Intelligent search service technology trends and industry outlook", Korea Communications Agency, PM Issue Report, No. 2, Issue. 4, May. 2013.
- [2] Oliver Oechslein, Thomas Hess, "Incorporating Social Networking Information in Recommender Systems: The Development of a Classification Framework", 26th Bled eConference eInnovations: Challenges and Impacts for Individuals, Organizations and Society, pp. 287-298, Jun. 2013.
- [3] Suyun Wei, Ning Ye, Shuo Zhang, Xia Huang, Jian Zhu, "Collaborative Filtering Recommendation Algorithm Based on Item Clustering and global Similarity", 2012 Fifth International Conference on Business Intelligence and Financial Engineering, pp. 69-72, Aug. 2012.
- [4] Tevfik Aytekin, Mahmut Ozge Karakaya, "Clustering-based diversity improvement in top-N recommendation", Journal of Intelligent Information Systems, Springer, pp. 1-18, Jun. 2013.
- [5] Yongli Ren, Gang Li, Wanlei Zhou, "A learning method for Top-N recommendation with incomplete data", Social Network Analysis and Mining, Springer, pp. 1-14, Feb. 2013.
- [6] Ernesto Diaz-Aviles, Lucas Drumond, Lars Schmidt-Thieme, Wolfgang Nejdl, "Real-time top-n recommendation in social streams", Proceedings of the sixth ACM conference on Recommender systems, pp. 59-66, Sep. 2012.
- [7] Michail Vlachos, Daniel Svonava, "Graph Embeddings for Movie Visualization and Recommendation", Proceedings First International Workshop on Interfaces for Recommender Systems, pp. 56-59, Sep. 2012.
- [8] Jason J. Jung, "Visualizing Recommendation Flow on Social Network", Journal of Universal Computer Science, Vol. 11, Issue. 11, pp. 1780-1791, Nov. 2005.
- [9] Inay Ha, Kyeong-Jin Oh, Myung-Duk Hong, Geun-Sik Jo, "Social filtering using social relationship for movie recommendation", The 4th international conference on Computational Collective Intelligence: technologies and applications, Vol. Part I, LNAI 7653, pp. 395-404, Nov. 2012.
- [10] Tim Berners-Lee, "Linked Data", <http://www.w3.org/DesignIssues/LinkedData.html>
- [11] Won Seok Oh, "Increase the value of the data linked data", Korean Semantic Web Conference, Dec. 2010.
- [12] Oktie Hassanzadeh, Mariano Consens, "Linked Movie Data Base", Proceedings of the WWW2009 Workshop on Linked Data on the Web, Apr. 2009.

- [13] Tommaso Di Noia, Roberto Mirizzi, Vito Claudio Ostuni, Davide Romito, Markus Zanker, "Linked Open Data to support Content-based Recommender Systems", 8th International Conference on Semantic Systems, Sep. 2012.
- [14] Roberto Mirizzi, Tommaso Di Noia, Azzurra Ragone, Vito Claudio Ostuni, Eugenio Di Sciascio, "Movie Recommendation with DBpedia", 3rd Italian Information Retrieval Workshop, Jan. 2012.
- [15] Tommaso Di Noia, Roberto Mirizzi, Vito Claudio Ostuni, Davide Romito, "Exploiting the Web of Data in Model-based Recommender Systems", 6th ACM Conference on Recommender System, pp. 253-256, Sep. 2012.
- [16] Federica Cena, Elisa Chiabrando, Andrea Crevola, Martina Deplano, Cristina Gena, and Francesco Osborne, "A Proposal for an Open Local Movie Recommender", 6th International Workshop on Personalized Access to Cultural Heritage, Jun. 2013.
- [17] Oktie Hassanzadeh, Mariano Consens, "Linked Movie Data Base", LOD Triplification Challenge at I-Semantics 2008, Apr. 2008.
- [18] Michail Vlachos, Daniel Svonava, "Graph Embeddings for Movie Visualization and Recommendation", Workshop on Interfaces for Recommender Systems 2012, in conjunction with the 6th ACM conference on Recommender Systems, pp. 56-59, Sep. 2012.
- [19] Daniel Svonava, Michail Vlachos, "Visualization and Recommendation of Similar Movies on the IMDB Movie Database.", Dec. 2010.
- [20] Svetlin Bostandjiev, John O'Donovan, Tobias Höllerer, "TasteWeights: A Visual Interactive Hybrid Recommender System", The 6th ACM conference on Recommender Systems, pp. 35-42, Sep. 2012.
- [21] John O'Donovan, Brynjar Gretarsson, Barry Smyth, "A Visual Interface for Social Information Filtering", International Conference on Computational Science and Engineering, Vol.4, pp. 74-81, Aug. 2009.
- [22] David Gotz, Zhen Wen, "Behavior-Driven Visualization Recommendation", The 14th international conference on Intelligent User Interfaces, pp. 315-324, Feb. 2009.
- [23] Katrien Verbert, Denis Para, Peter Brusilovsky, Erik Duval, "Visualizing Recommendations to Support Exploration, Transparency and Controllability", The 18th international conference on Intelligent User Interfaces, pp. 351-362, Mar. 2013.
- [24] Herlocker, J. L., Konstan, J. A., Terveen, L. G., Riedl, J. T., "Evaluating Collaborative Filtering Recommender Systems", ACM Transactions on Information Systems, Vol. 22(1), pp. 5-53, Jan. 2004.
- [25] Kurt Bollacker, Colin Evans, Praveen Paritosh, Tim Sturge, Jamie Taylor, "Freebase: A Collaboratively Created Graph Database For Structuring Human Knowledge", Proceedings of the ACM SIGMOD International Conference on Management of Data, pp. 1247-1249, Jun. 2008.

저 자 소 개



홍 명 덕
 2008 : 서울디지털대학교
 컴퓨터공학과 공학사
 2011 : 인하대학교
 컴퓨터정보공학과 공학석사
 2011~현재 : 인하대학교
 컴퓨터정보공학과
 박사과정
 관심분야 : 추천 시스템, 시맨틱 웹,
 군집 지능, 메타휴리스틱
 Email : hmdgo@eslab.inha.ac.kr



하 인 애
 2005 : 수원대학교
 컴퓨터과학과 공학사
 2007 : 인하대학교
 컴퓨터정보공학과 공학석사
 2007~현재 : 인하대학교
 컴퓨터정보공학과
 박사과정
 관심분야 : 개인화, 시맨틱웹,
 소셜네트워크, 추천 시스템
 Email : inay@eslab.inha.ac.kr



조 근 식
 1982 : 인하대학교
 전자계산학과 공학사
 1985 : Queens Colleg/CUNY M.A.
 컴퓨터공학과 공학석사
 2004 : City University of New York
 컴퓨터공학과 공학박사
 1991~현재 : 인하대학교
 컴퓨터정보공학과 교수
 관심분야 : 인공지능, 시맨틱 웹,
 전문가시스템,
 지능형에이전트시스템
 Email : gsjo@inha.ac.kr