

사회적 관계 기반 매쉬업 콘텐츠 추천 기법

서현^o 윤효석 우운택

광주과학기술원 U-VR 연구실

hseo@gist.ac.kr hyoon@gist.ac.kr wwoo@gist.ac.kr

Mashup Content Recommendation based on Social Relationship

Hyeon Seo^o Hyoseok Yoon Woontack Woo

Gwangju Institute of Science and Technology, U-VR Lab

1. 서 론

최근 다양한 소셜 미디어를 조합하여 새로운 형태의 콘텐츠, 서비스로 만드는 매쉬업(mashup)이 주목을 받고 있다. 매쉬업 콘텐츠는 일종의 소셜 미디어로써 사용자들이 여러 형태로 생성 할 수 있으며, 생성된 매쉬업 콘텐츠에 특정 키워드와 태그를 등록할 수 있다. 이렇게 생성된 매쉬업 콘텐츠에 대해 사용자들은 추천 활동을 통해 매쉬업 콘텐츠를 평가하고 그들의 의견, 경험, 생각을 공유 할 수 있다. 다 수의 사용자에게서 다양한 형태로 생성된 매쉬업 콘텐츠의 수가 많아질수록 개별 사용자에게 적합한 매쉬업 콘텐츠를 분별하여 제공하기는 더욱 어려워진다. 본 논문에서는 이런 문제점을 해결하기 위해 사회적 관계에 기반한 새로운 추천 기법을 제안하고 기존의 추천 기법과 비교한다. 특히, 새롭게 제안하는 기법은 매쉬업 콘텐츠를 이용하는 불특정 다수 사용자의 선호도 정보와 사용자와 높은 유사도/신뢰도를 가지는 유사 사용자의 선호도 정보를 활용한다. 사회적 관계 기반 추천 기법의 유용성을 검증하기 위해 시뮬레이션을 통해 생성한 사용자 프로파일과 매쉬업 콘텐츠를 통해 알고리즘 측면에서 추천 기법을 비교 분석한다.

2. 사회적 관계 기반 추천

본 논문에서는 소셜 미디어의 사회적 관계에 기반 한 기존의 추천 기법인 인기도 기반 추천 기법(Popularity-based, POP), 협업 필터링 기법(Collaborative Filtering, CF)과 새로운 추천 기법(Social Hierarchy, SH)을 제안하고 비교한다. POP은 불특정 다수라는 사회적 관계를 통해 많은 사용자가 추천한 매쉬업 콘텐츠를 일반적으로 의미 있는 또는 인기 있는 콘텐츠로 간주하여 높은 추천수의 매쉬업 콘텐츠를 사용자에게 추천한다. CF는 피어슨 상관계수를 이용하여 사용자들 간의 유사도를 계산하고 [1] 뷰어와 다른 사용자 간의 추천 패턴을 통해 계산된 유사도에 기반 한 사회적 관계를 활용한다. 이를 통해 매쉬업 콘텐츠에 대한 예상 선호도를 예측하고 높은 예상 선호도를 가진 매쉬업 콘텐츠를 사용자에게 추천한다.

CF에서는 두 사용자간의 유사도 비교를 통해 사용자의 가중치를 측정하였다면 SH 추천 기법은 사용자의 직접적/간접적 사회적 관계를 통해 사용자에게 가중치를 주고 이를 이용해 매쉬업 콘텐츠의 예상 선호도를 예측한다. SH기법은 사용자의 가중치와 추천 활동을 통해 매쉬업 콘텐츠의 선호도를 구하고 예상 선호도에 따른 상위 k개의 매쉬업 콘텐츠를 사용자에게 추천한다.

SH 기법에서 고려하는 사회적 관계란 사용자 프로파일에 명시된 사용자의 가족, 친구, 기타 그룹의 관계와 유사 선호도를 가진 사용자들의 그룹이다. 현재 단계에서는 사용자 프로파일을 통해 명시적으로 알 수 있는 사회적 관계[2]를 사용하고 있다. 사용자의 u 의 가중치 $w(u)$ 는 다음과 같다.

본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 지원사업 (NIPA-2010-C1090-1011-0008)과 문화체육관광부 및 한국콘텐츠진흥원의 2010년도 문화콘텐츠산업기술지원사업의 연구결과로 수행되었음

$$w(u) = \alpha + \beta \text{sim}(\text{viewer}, u) + \gamma \text{SR}(\text{viewer}, u), \quad \text{where } \alpha + \beta + \gamma = 1$$

위 식에서 α 는 각 사용자에게 주어지는 기본 값이며 $\text{sim}(\text{viewer}, u)$ 를 통해 뷰어와의 유사도 정도에 따라 사용자의 가중치를 부과한다. $\text{SR}(\text{viewer}, u)$ 는 사용자와의 직접적 사회적 관계에 따른 수식이다. 이는 다음과 같다.

$$\text{sim}(\text{viewer}, u) = |A \cap B| / |A \cup B|$$

where A = the set of viewer's interest keywords
 B = the set of u's interest keywords

$$\text{SR}(\text{viewer}, u) = \begin{cases} 1 & \text{if } u \text{ is a family member} \\ 0.5 & \text{if } u \text{ is a friend} \\ 0.3 & \text{if } u \text{ is a group member} \end{cases}$$

$\text{sim}(\text{viewer}, u)$ 는 다른 사용자들 간의 유사도를 나타내며 Jaccard Index를 이용해 이를 계산하였다. $\text{SR}(\text{viewer}, u)$ 는 뷰어의 직접적 사회적 관계를 의미한다. 위의 과정을 통해 구한 사용자의 가중치와 추천활동을 통해 매쉬업 콘텐츠에 대한 선호도를 구한다. 이때 구글의 단순화된 페이지 랭크 알고리즘[3]을 적용하여 매쉬업 콘텐츠를 추천하는 사용자들의 가중치에 따라 개인화되고 동시에 많은 이들이 선호하는 매쉬업 콘텐츠를 추천하고자 한다. i 번째 매쉬업 콘텐츠 MC_i 의 선호도는 다음과 같다.

$$RV(MC_i) = \sum_{u \in U} \frac{w(u)}{L(u)}$$

MC_i 를 추천한 사용자들의 집합 U 의 사용자 u 에 대해 $L(u)$ 는 사용자 u 가 투표한 총 수이며 $w(u)$ 는 사용자의 가중치이다. 이를 통해 불특정 다수의 사용자들도 선호하지만 동시에 뷰어와 사회적 관계가 형성되어 있는 사용자들의 선호도가 반영된 추천 결과를 얻을 수 있다.

본 논문에서 제시한 사회적 관계 기반 3가지 추천 기법인 추천수 기반(POP), 협업 필터링 기반(CF), 사회적 계층 기반(SH)과 내용 기반 추천 기법(Content-based Filtering, CB)을 가상 시뮬레이션 데이터를 사용하여 비교하였다. 각 알고리즘을 통해 추천된 k 개의 매쉬업 콘텐츠를 비교 평가하기 위하여 정보검색에서 사용되는 Precision, Recall, F-Measure와 Fall-Out을 사용하였다. 이 때 적합한 매쉬업 콘텐츠(RMC: Relevant Mashup Content)의 기준으로는 1) 추천된 매쉬업 콘텐츠가 사용자 관심사 키워드를 포함 2) 전체 1,000개의 매쉬업 콘텐츠 중 추천수가 높은 상위 k 개의 매쉬업 콘텐츠로 정의한다. 적합한 매쉬업 콘텐츠의 정의에 의해 POP 알고리즘의 성능이 우수하게 나타난다. POP을 기준 알고리즘으로 비교하면 SH 알고리즘의 성능이 CB, CF보다 높은 성능을 보여준다. 또한 랭킹에 포함된 사회적으로 적합한 매쉬업 콘텐츠(SMC: Socially-relevant Mashup Content) 수를 통하여 네 가지 알고리즘을 비교하였을 때 CB는 사회적으로 적합한 매쉬업 콘텐츠를 추천하기에는 적합하지 않으나 POP, CF, SH는 평균적으로 높은 수의 사회적으로 적합한 매쉬업 콘텐츠를 랭킹에 포함한다.

3. 결 론

최근에 대두하고 있는 소셜 네트워크 서비스(SNS)와 소셜 미디어로 인해 온라인에서도 사용자의 인맥, 사회적 관계가 중요하게 고려되고 있다. 본 논문에서는 다 수의 다양한 소셜 미디어로 생성 될 수 있는 매쉬업 콘텐츠를 사용자의 사회적 관계기반으로 추천할 수 있는 여러 기법을 제시하고 시뮬레이션 데이터를 통해 알고리즘 측면에서의 특성을 살펴보았다. 사회적 관계를 사용하지 않는 추천 기법인 내용 기반 기법에 비해 불특정 다수의 추천을 사용하는 POP, 협업 필터링, 사회적 계층 방법은 평균적으로 더 많은 수의 사회적으로 적합한 매쉬업 콘텐츠를 랭킹에 포함하는 경향을 살펴볼 수 있었다. 추후 계획으로는 알고리즘 측면에서 살펴본 사회적 관계기반의 매쉬업 콘텐츠 추천 기법을 확장하여 소셜 네트워크 서비스 및 소셜 미디어의 실제 데이터를 통해 적용하고 검증할 예정이다.

참고문헌

[1] Xiaoyuan Su and Taghi M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques," Advances in Artificial Intelligence, vol. 2009, Article ID 421425, 19 pages, 2009.
 [2] Kristina Lerman, "Social Information Processing in News Aggregation," IEEE Internet Computing, pp. 16-28, 2007.
 [3] Taher H. Haveliwala "Topic-Sensitive PageRank : A Context-Sensitive Ranking Algorithm for Web Search," IEEE Transactions on Knowledge and Data Engineering, vol. 15, no. 4, pp. 784-796, 2003.