

소셜 네트워크의 클러스터 구조를 활용한 콘텐츠 추천

신윤섭¹ 서동환⁰² 황지영^{2*}

¹성균관대학교 전자전기컴퓨터공학과

²성균관대학교 컴퓨터공학과

{skcy2004, donghwan79, jjwhang}@skku.edu

Content Recommendation using the Clustering Structure of Social Networks

Yunseob Shin¹ Donghwan Seo⁰² Joyce Jiyoung Whang^{2*}

¹Department of Electrical and Computer Science, Sungkunkwan University

²Department of Computer Science and Engineering, Sungkyunkwan University

요 약

소셜 네트워크가 광범위하게 형성되면서 웹상에서도 사회적 관계가 중요하게 되었다. 그러나 현재 널리 사용되는 협업 필터링과 내용기반 추천 시스템은 사회적 관계를 반영할 수 없으며, 사용자의 행동 기록이 적을 경우 추천 정확도가 현저히 낮아지게 된다는 문제가 있다. 따라서 본 연구에서는 소셜 네트워크 그래프 기반의 클러스터를 활용해 사용자의 사회적 관계와 현실 세계에서의 선호도를 웹상에서 반영하여, 사용자에게 적절한 콘텐츠를 추천하는 방법을 제안한다. 사용자와 콘텐츠와의 관계가 웹 페이지의 링크 관계와 유사하다는 점에 착안해, HITS 알고리즘을 응용한 클러스터 기반 콘텐츠 추천 평가 시스템을 제안했으며 기존의 방법론과 차이가 있음을 확인했다.

1. 서 론

소셜 네트워크(social networks)는 소셜 그래프(Social Graph)로 표현할 수 있다. 소셜 네트워크상에서 활동하는 각각의 개인들을 노드(node)로, 개인들 간의 관계를 링크(tie) 혹은 간선(edge)으로 표현하는 그래프로 모델링 할 수 있다. 소셜 네트워크 상에서 사용자들은 의도적으로 관계를 맺거나 비슷한 특징을 공유하기 때문에 그래프 상에서 간선(edge)이 밀도 있게 형성되는 특정한 그룹들이 존재하는데, 이러한 그룹들을 클러스터(cluster)라고 부른다. 소셜 그래프 내의 클러스터들을 분류하고 각각의 노드의 속성을 파악할 수 있다면, 이를 활용해 클러스터 내의 타 사용자가 어떤 유형의 콘텐츠를 선호할지 예측할 수 있다.

특정 사용자의 선호도는 해당 클러스터 내의 다른 사용자, 즉 간선으로 연결된 노드(node)의 특정 콘텐츠에 대한 성향을 분석하여 파악할 수 있다. 페이스북과 같은 소셜 네트워크 서비스 상에서 사용자들은 특정 분야의 콘텐츠를 전문적으로 제공하는 페이지(page)를 팔로우(link)하여 주기적으로 수용하기 때문에 이를 근거로 노드의 선호 속성을 알아 낼 수 있다. 클러스터 내의 노드들과 페이지 간의 링크 관계를 파악하게 되면,

해당 클러스터 내에서 어떤 콘텐츠가 사용자에게 높은 만족도를 줄 수 있을지 점수를 계산해 순위(rank)를 도출할 수 있다.

특히 소셜 네트워크 상에서 사용자들은 오프라인 관계를 기반으로 활동하는 경우가 많기 때문에, 소셜 그래프 또한 실제 현실 세계에서의 관계를 중심으로 그래프 간선이 밀집되어 있다. 즉, 소셜 네트워크의 클러스터를 활용한 추천 시스템은 현실 세계에서의 관계를 적극적으로 반영한 알고리즘이며, 기존의 추천 알고리즘이 온라인 내에서의 행동프로필(profile)만 반영한 것과 가장 큰 차이를 갖는 부분이다. 이것이 중요한 이유는 온라인에서의 콘텐츠 소비가 사용자의 실제 현실에 소속된 또래 집단, 성별, 계층, 직업 등에 의해 영향을 크게 받기 때문이다. 또한 소셜 네트워크 상에서는 사용자가 관심 콘텐츠 페이지나 타 사용자에게 의도적으로 링크를 할 수 있기 때문에 기존의 추천 알고리즘보다 사용자의 선호 속성을 더 빠르고 정확하게 반영할 수 있다.

본 논문에서는 소셜 네트워크의 클러스터를 활용한 콘텐츠 추천을 제안한다. 전체 사용자 간 연결 그래프에 Graclus 알고리즘을 적용해 클러스터링을 수행했고[4],

* 교신저자(Corresponding author)

이 논문은 2017년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(과제번호: 2016R1D1A1B03934766)

추천 시 콘텐츠 우선순위를 판별하기 위해 HITS 알고리즘을 응용한 Local Authorities Score 방법을 개발했다. 인접 행렬(adjacency matrix)로 표현된, 클러스터와 콘텐츠 페이지 사이의 링크 구조 데이터에 제안하는 방법론을 적용하면 클러스터에 기반(cluster-based)한 콘텐츠들의 추천 순위(score rank)를 도출할 수 있다.

2. 관련연구

모델 및 메모리 기반 협업 필터링(Collaborative Filtering)은 웹상에서 사용자의 프로파일이나 클릭 기록 등을 수집해 비슷한 성향의 사용자끼리 클러스터를 형성하고 이를 바탕으로 콘텐츠를 추천한다[1]. 그러나 사용자의 활동과 프로파일 적을 경우 유의미한 추천이 어려우며, 사용자 간 상호작용과 사회적 관계를 추천 평가 항목에 반영할 수 없다는 문제가 있다. 협업 필터링의 단점을 보완하기 위해 제안된 확률론적 행렬 인수분해 기법(Probabilistic Matrix Factorization)을 활용한 소셜 네트워크 기반 추천 알고리즘은 사용자의 행동 기록이 없는 상황에서도 소셜 그래프를 활용해 콘텐츠를 추천해줄 수 있으며 더 많은 데이터(datasets)를 다룰 수 있다[2]. 이 기법에서는 사용자-아이템 행렬(user-item matrix)과 소셜 그래프를 활용하여 해당 아이템에 대해 행동 기록이 없던 사용자의 만족도를 예측한다.

HITS 알고리즘은 하이퍼링크 구조를 활용해 검색쿼리에 대해 가장 권위 있는(authority) 웹페이지를 추출해 낸다[3]. 다른 페이지들로부터 링크가 많이 되어 있는 페이지일수록 높은 authority score를 갖고, 이 페이지들에 링크를 많이 하고 있을수록 높은 hub score를 갖는다. 각 가중치를 반복적으로 연산하면 수렴된 authority score를 구할 수 있으며, 쿼리에 대해 어떤 페이지가 가장 정확한지 순위(rank)를 얻을 수 있다. 소셜 네트워크 상에서 사용자와 콘텐츠 페이지 간의 관계가 이와 유사한 구조를 가지고 있기 때문에 이를 응용해 추천 순위를 얻는 시스템을 소셜 그래프에 적용할 수 있다.

즉, 많은 친구들(related-user)이 소셜 네트워크 상에서 팔로우하고 있는 콘텐츠 페이지일수록 권위(authority score)가 높기 때문에 사용자에게 해당 콘텐츠를 추천했을 시 만족도가 높을 가능성이 크며, 클러스터 관점에서의 링크를 고려했기 때문에 사용자의 사회 관계가 직접적으로 반영되어 콘텐츠에 대한 만족도가 더 높을 것이다.

3. 클러스터 기반 추천 알고리즘 적용

3.1 소셜 그래프 및 페이지링크 행렬 생성

128,821개의 노드와 1,367,333개의 콘텐츠 페이지들을 이용해 두 노드 사이에 간선 존재 여부를 확인하는 사용자-사용자 행렬과, 노드와 페이지 사이의 링크 관계를 확인하는 사용자-페이지 행렬을 만든다.

3.2 클러스터링 및 authority score 계산

3.2.1 Graclus

Graclus 알고리즘은 고유벡터 계산 없이 주어진 무방향 그래프에 대해 normalized-cut을 최적화하는 그래프 클러스터링 알고리즘이다[4].

3.2.2 Centrality

Centrality는 그래프에서 해당 노드의 중요도를 나타내는 점수이다. Centrality를 측정하는 방법으로는 PageRank, hubs and authorities 등이 있다. 이 논문에서는 페이지와 사용자라는 각자 다른 속성을 가진 노드들로 이루어진 그래프이기 때문에 hubs and authorities를 사용했다. 사용자는 hub score, 페이지는 authority score를 갖는다. 이를 통해 권위 있는 페이지를 추출하고 순위를 도출할 수 있도록 페이지의 authority score를 계산한다.

3.3 추천 시스템 적용

클러스터 성질이 가미되지 않은 Global Authority Score와 클러스터 성질을 가미한 Local Authority Score를 구하여 점수와 순위를 비교한다.

3.3.1 Global Authority Score

- (1) 모든 콘텐츠 페이지의 authority score를 계산한다.
- (2) Graclus를 적용해 사용자 노드들을 클러스터링 한다.
- (3) 각 클러스터와 링크된 페이지들을 대상으로 순위를 비교하기 위해 (1) 단계에서 얻은 Global Authority Score를 각 클러스터 별로 정규화(normalization)한다.
- (4) 각 클러스터에 링크된 페이지들의 추천 순위를 얻는다.

3.3.2 Local Authority Score

- (1) Graclus를 적용해 각 노드들을 클러스터링 한다.
- (2) 각 클러스터와 연결된 콘텐츠 페이지들을 대상으로 authority score를 계산한다.
- (3) 각 클러스터에 링크된 페이지들의 추천 순위를 얻는다.

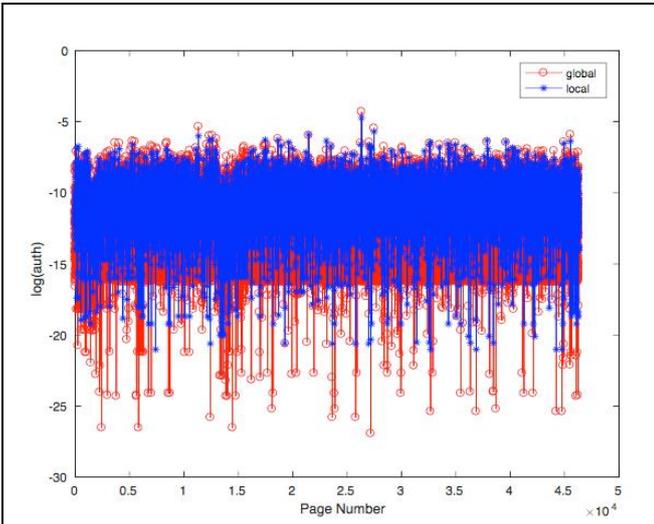
Global Authority Score와 Local Authority Score는 authority score를 구하는 시점에서 차이를 갖는다. Global Authority Score는 클러스터링을 하기 전에 이미

구한 authority score를 바탕으로 클러스터 내의 순위를 도출하고, Local Authority Score는 클러스터링 이후에 authority score를 구하여 각 클러스터별 순위를 정한다.

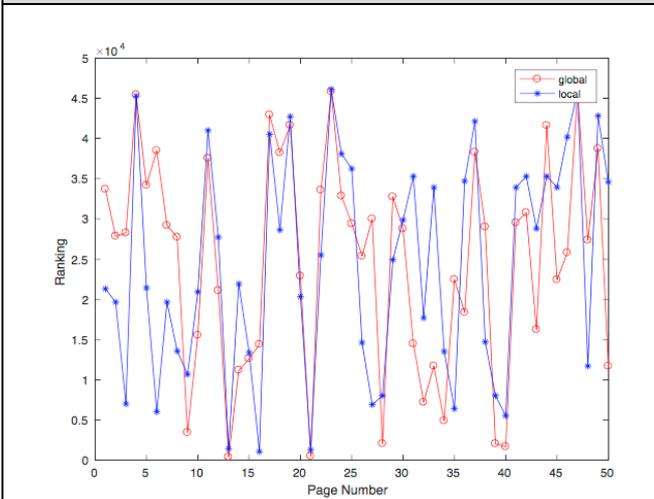
4. 실험 결과

실험에서 Global Authority Score와 Local Authority Score를 소셜 그래프에 적용하여 결과를 비교했다.

[표 1]은 1번, 12번, 48번 세 개 클러스터에서 Global Authority Score와 Local Authority Score별로 상위 10개의 페이지 번호를 의미한다. 추천순위 상위 10개 페이지의 authority score는 1위를 제외하고 대부분 겹치지 않게 나타난다. 이를 통해 클러스터 적용 여부에 따라 페이지의 추천 순위에 명확한 차이가 있음을 확인했다.



(a) 12번 클러스터의 Page-authority 그래프



(b) 12번 클러스터의 Page-authority 순위 그래프 일부

<그림 1> 12번 클러스터의 page-authority 및 ranking 그래프

<그림 1-a>는 64개의 클러스터 중 12번 클러스터에서 전체 페이지들의 Global Authority Score와 Local Authority Score를 그래프로 표현한 것이다. <그림 1-b>는 그 중에서 50개 페이지를 임의로 선정하여 페이지별 순위를 나타냈다. 같은 클러스터에서 Global과 Local의 관점에 따라 페이지의 authority score와 추천 순위가 대부분 다르게 나타나는 것을 알 수 있다. 따라서 클러스터 적용 여부에 따라 score 및 추천 순위에 명확한 차이가 있음을 확인했다.

	Cluster 12		Cluster 1		Cluster 48	
Rank	Global	Local	Global	Local	Global	Local
1	764828	764828	764828	764828	764828	764828
2	289455	799307	289455	615125	289455	289455
3	799307	615125	799307	799307	799307	330085
4	1341359	289455	1341359	289455	330085	187977
5	615125	320467	615125	314719	123143	314923
6	320467	491398	320467	528082	86369	123143
7	314719	314719	314719	320467	212805	177083
8	330085	528082	330085	1006786	787834	218916
9	528082	1115172	528082	1341359	176745	275051
10	118336	665352	118336	593910	153871	277862

[표 1] 12, 1, 48번 클러스터의 authorities rank 상위 10개 비교

5. 결론 및 향후 연구

본 논문에서는 기존 추천 알고리즘의 단점을 보완하기 위해 소셜 그래프의 클러스터를 활용하는 방법론을 제안했다. Global Authority Score와 Local Authority Score의 결과가 서로 다르게 나타나는 것을 확인했고, 추천 시스템을 설계할 때 전통적인 hubs and authorities의 개념을 활용한 Global Authority Score뿐만 아니라 클러스터의 성질을 활용하는 Local Authority Score도 적극적으로 고려해야 한다는 것을 확인했다. 추후 추천 알고리즘의 성능을 개선하여 궁극적으로는 거대 소셜 네트워크 등 빅데이터를 활용할 수 있는 정교한 추천 알고리즘으로 개선하는 것을 목표로 하고 있다.

참 고 문 헌

[1] Breese et al. "Empirical analysis of predictive algorithms for collaborative filtering." *UAI*, 1998.

[2] Ma. Hao et al. "Social Recommendation using probabilistic matrix factorization." *CIKM*, 2008.

[3] Kleinbera. Jon M. "Authoritative sources in a hyperlinked environment." *Journal of the ACM*, 1999.

[4] Dhillon et al. "Weighted graph cuts without eigenvectors a multilevel approach." *PAMI*, 2007.