**논문 요약**

**날짜: 2020.11.25**

**이름: 전우진**

|  |  |
| --- | --- |
| 논문제목 | Friendship Prediction and Homophily in Social Media |
| 논문 요약 | 온라인 소셜 네트워킹 시스템은 콘텐츠 구성 및 공유를 위한 온라인 시스템과 함께 기반 기술 플랫폼을 통해 사용자 커뮤니티의 인지, 행동 및 사회적 측면을 혼합한다. 그 결과로 나타나는 "생태계"는 개인의 상호작용에서 작용하는 다양한 과정을 채굴하고 조사할 수 있는 새로운 가능성을 제공하며, 사용자가 공유하는 정보와 어떻게 관련되는지 연구할 수 있다. 태그를 기반으로 한 사회적 주석은 사용자, 관심 항목(리소스), 메타데이터(태그)의 3방향 관계를 외부화하기 때문에 연구에 가치가 있다. 본 연구에서는 세 가지 다른 온라인 소셜 시스템을 고려한다. Flickr, Last.fm, 그리고 aNobiii. 이러한 시스템에서 사용자는 자원(각각 사진, 노래/예술가, 책)을 공유하고 소셜 네트워크를 형성하며, 태그 항목에 태그를 지정하여 일반적으로 민중학이라고 불리는 데이터의 사회적 분류를 생산한다.  이전 연구에서는 온라인 소셜 네트워크에서의 근접성과 국소적 유사성 사이의 상관관계를 탐구했으며, 태그 정보에 근거한 사용자들 사이의 유사성이 사회적 유대를 정확하게 예측할 수 있는 정도를 분석했다.  2004년 1월부터 2006년 1월 사이에 플리커에 업로드된 사진에 대한 태그 정보를 API 메소드(flickr.com/api))로 수집했다. 전체 탐색은 2007년 상반기 동안 수행되었다. 분석은 태그, 그룹, 연락처 정보를 가지고 있는 약 13만 명의 사용자들의 네트워크에 초점을 맞출 것이다.  Last. fm에서 각 사용자는 양쪽 엔드포인트의 동의 하에 설정된 리디렉션되지 않은 링크를 통해 친구와 연결된다. 크롤링은 2010년 1월에 수행됐다.  아노비(anobii.com) 사용자들은 49개 언어로 작성된 약 2000만 종의 다양한 출판물의 메타데이터(출판 연도, 저자 등)를 수록한 공공 아노비 도서 데이터베이스에서 선택한 제목으로 디지털 도서 컬렉션을 채운다. 각각의 개인 도서 모음집은 이용자가 읽고 있거나 이미 읽은 책들의 집합체인 도서관과 가까운 장래에 읽고 싶은 책들을 나열하는 위시리스트로 구분된다. 우리는 2009년 12월에 무작위 사용자들로 시작해 포워드 BFS 패션으로 소셜 링크를 따라가는 aNobii 네트워크를 탐색했다.  우리는 소셜 네트워크의 모든 거대하고, 강하게 연결된 컴포넌트와 아웃 컴포넌트를 총 86,800명의 사용자를 대상으로 조사했다.  소셜 네트워크는 사용자에 의해 명백하게 구축되기 때문에, 우리는 또한 우정 관계의 수를 우리가 고려하는 각 민간의 활동 척도로 생각할 수 있다.  링크가 연출될 때, 우리는 out-degree를 activity의 대리지표로 간주하는 반면, in-degree는 인기도를 측정한다.  먼저 개별 사용자 특징과 사회적 상호작용 사이의 상호작용을 점검했다.  우리는 사회적 연결성과 태깅과 그룹 참여와 같은 명시적 사용자 활동의 강도 사이의 강한 상관관계를 관찰했다.  이웃들 사이의 다양한 혼합 패턴은 또한 검사된 모든 특징에 대해서도 발견될 수 있다.  우리는 사용자 쌍 간의 국소적 유사성과 사회적 그래프의 최단거리 사이의 상관관계를 보여주는데, 이는 순전히 통계적 분류성과 실제 동음이의성을 구별하기 위해 구축된 null 모델을 사용한다.  우리는 조사된 모든 특징에 대해 이웃 간에 명확한 주제적 정렬 경향을 발견했다.  동음이의에 대해 얻은 결과는 Last.fm과 aNobii의 사용자 프로필 특징을 바탕으로 소셜 링크의 존재를 예측할 수 있는 가능성을 조사하도록 이끌었다.  우리는 몇 가지 유사성 지표를 사용하여 이용 가능한 각 기능(태그, 태그된 항목, 그룹 및 라이브러리 항목)에 대한 사용자 유사성 네트워크를 구축했다.  소셜 링크 예측을 위해 사용자 쌍의 유사도 순위를 사용했으며, ROC 곡선을 사용하여 예측 정확도를 평가하였다.  결과는 모든 사용자 프로파일 기능이 상당한 예측 능력을 가지고 있음을 보여준다.  민중학에서 파생된 특징에 대해서는, 우리가 민중학 공간에서 물체를 비교하기 위해 도입한 유사성 메트릭, 즉 MIP 유사성이 가장 정확한 결과를 산출했다.  더욱이, 3차원 민중학 공간에서 민중학 특징을 추출하는 데 사용되는 집적 기술은 예측 결과와 매우 관련이 있다.  전체적으로 도서관 기능, 즉 사용자 아이템(아노비이의 책과 라스트의 아티스트)의 모음입니다.  fm)은 예측 가능한 항목으로 판명되었고, 사용자가 적은 수의 라이브러리 항목을 가지고 있을 때에도 매우 정확했다.  마지막으로, 종단 피쳐와 기계 학습 접근방식을 통해 네트워크 기반 피쳐를 결합하면 링크 예측 정확도가 향상된다.  구어, 국적 등 혼동적인 측면이 예측 과제에 미치는 영향도 연구했다.  서로 다른 클러스터에 속한 사용자들은 매우 다른 주제적 관심사를 가지고 있으며 사회적 유대관계를 거의 설정하지 않기 때문에 언어별로 강하게 군집화된 소셜 네트워크에서는 링크 예측 과제가 더 쉽다는 것을 보여주었다.  이 때문에 실제 소셜 네트워크 데이터에 대한 특징 기반 링크 예측 알고리즘을 테스트하기 전에 언어 공동체 구조에 대한 사전 분석을 실시할 것을 제안한다. |
| 논문 의의 | 태그를 기반으로 한 사회적 주석은 사용자, 관심 항목(리소스), 메타데이터(태그)의 3방향 관계를 외부화하기 때문에 연구에 가치가 있고, 세 가지 다른 온라인 소셜 시스템을 함께 고려해 다양한 사용자 활동의 측면을 고려할 수 있다. |
| 논문의 한계점 및 기타 | 현재의 연구 결과는 많은 가능한 미래 방향을 제시한다.  첫째, 동음이의 인과관계를 상세히 탐구할 만하다.  보다 일반적으로 사회 환경의 종적 스냅숏을 고려할 때 연결성과 유사성 사이의 상호 작용의 진화적 패턴을 연구할 수 있다.  또한 임시 데이터 세트를 사용하면 예측된 링크가 가까운 미래에 실제로 나타나는지 여부를 확인함으로써 링크 예측 접근법의 유효성을 더 잘 테스트할 수 있다. |