**논문 요약**

**날짜: 2020.11.25**

**이름: 전우진**

|  |  |
| --- | --- |
| 논문제목 | Link-Prediction Problem for Social Networks |
| 논문 요약 | 크고 복잡한 네트워크와 그 속성에 대한 최근 급증하는 연구의 일부로서, 상당한 관심이 소셜 네트워크의 컴퓨터 분석(노드가 사회적 맥락에 내재된 사람 또는 다른 실체를 나타내며, 가장자리가 상호작용, 협업 또는 영향력을 나타내는 구조)에 집중되었다.  우리는 소셜 네트워크 진화의 기초가 되는 기본적인 컴퓨터 문제인 링크-프레전스 문제를 정의하고 연구한다. 시간 t에서 소셜 네트워크의 스냅샷을 제공하면, 우리는 시간 t에서 주어진 미래 시간 t까지의 간격 동안 네트워크에 추가될 가장자리를 정확하게 예측하려고 한다.  링크 예측을 위한 효과적인 방법은 조직 내에서 아직 확인되지 않은 유망한 상호작용이나 협력을 제안하기 위한 소셜 네트워크를 분석하는 데 사용될 수 있다. 다른 맥락에서 최근 안보 분야 연구가 '역할'을 강조하기 시작했다.  데이터는 [www.arxiv.org에서](http://www.arxiv.org에서) 5가지의 coauthorship networks를 구성한다. 우리는 훈련 간격을 1994년부터 1996년까지의 3년, 시험 간격은 1997년부터 1999년까지의 3년으로 정의한다.  링크 예측을 위한 일련의 방법을 조사한다. 모든 방법은 입력 그래프 Gcollab을 기준으로 연결 중량 점수(x, y)를 x, y , 노드의 쌍에 할당하고 점수(x, y)의 감소 순서로 순위 목록을 생성한다.  그러므로 그것들은 네트워크 토폴로지에 상대적인 노드 x와 y 사이의 근접성 또는 "유사성"의 측정치를 계산하는 것으로 볼 수 있다.  첫 번째 절에서 논의한 바와 같이, 네트워크 범위 밖의 이유로 많은 협업이 형성(또는 형성되지 않음)되므로, 예측 변수의 raw performance는 상대적으로 낮다.  예측 변수의 품질을 보다 의미 있게 표현하기 위해, 우리는 무작위 예측 변수를 기준선으로 사용하며, 이 예측 변수는 훈련 간격에 협력하지 않은 무작위로 선택된 저자의 쌍을 단순히 예측한다. |
| 논문 의의 | 소셜 네트워크에서 새로운 상호작용이 가까운 미래에 일어날지 유추할 수 있는지를 링크 예측 문제로 바꾸고 네트워크상에서 노드의 「근접성」을 분석하는 대책에 근거해 예측을 연계하는 접근법을 개발한다. 대규모 공동저작권 네트워크에 대한 실험은 향후 상호작용에 관한 정보를 네트워크 토폴로지에서만 추출할 수 있으며, 노드 근접성을 검출하기 위한 상당히 미묘한 조치가 보다 직접적인 조치를 능가할 수 있음을 시사한다. |
| 논문의 한계점 및 기타 | 우리가 논의한 예측 변수들이 합리적으로 잘 수행되지만, 최고(gr-qc 상의 Katz 군집화)조차도 예측의 약 16%에 대해서만 정확하다.  이 과제에 대한 성과 개선의 여지가 분명히 많으며, 훈련 데이터의 정보를 더 잘 활용할 수 있는 방법을 찾는 것은 흥미로운 개방형 질문이다.  또 다른 문제는 매우 큰 네트워크에서 근접 기반 방법의 효율성을 개선하는 것이다. 노드 대 노드 거리 분포에 대한 빠른 알고리즘은 하나의 접근법이 될 수 있다. Gcollab 그래프는 데이터의 손실 표현이다. 우리는 또한 초당적 협력 그래프인 Bcollab을 고려할 수 있다. 여기에는 모든 저자와 기사에 대한 꼭지점, 그리고 각 기사를 각 저자와 연결하는 가장자리가 포함된다.  초당적 그래프는 Gcollab보다 더 많은 정보를 포함하고 있으므로 예측 변수들이 이를 사용하여 성능을 향상시킬 수 있기를 희망할 수 있다.  마찬가지로, 우리의 실험은 모든 훈련 기간의 협력을 동등하게 취급한다.  아마도 최근의 협업을 이전의 협력보다 더 중요하게 다루면서 성과를 향상시킬 수 있을 것이다.  또한, 예를 들어, 훈련 세트를 임시 세그먼트로 나누고, 시작 부분에 훈련 베타, 그리고 최종 예측을 위해 훈련 세트의 끝을 사용함으로써, Katz 예측 변수의 매개변수를 조정할 수 있다.  또한 각 과학자의 특정 연구 영역 또는 지리적 위치를 식별하기 위해 기사의 제목이나 저자의 기관 제휴와 같은 추가 정보를 사용하고, 그 다음 협업을 예측하기 위해 영역/로케이션들을 사용할 수 있다. |