**논문 요약**

**날짜: 2020.11.25**

**이름: 전우진**

|  |  |
| --- | --- |
| 논문제목 | Link Prediction with Signed Latent Factors in Signed Social Networks |
| 논문 요약 | 서명된 소셜 네트워크는 연결된 노드 쌍(즉, 사회적 상대성 또는 링크)을 긍정, 부정 또는 중립으로 표시할 수 있기 때문에 서명되지 않은 소셜 네트워크보다 실제 소셜 사이트의 복잡한 관계를 더 잘 반영한다.  또한, 연결되지 않은 두 개의 노드는 사회적 관계의 부재를 나타낸다. 예를 들어, 선거 네트워크에서, 유권자는 긍정, 부정 또는 중립적인 표를 던지거나 아예 투표하지 않기로 선택할 수 있다.  이러한 종류의 네트워크 구조에 대한 많은 원시적 용도를 감안할 때, 본 논문은 서명된 소셜 네트워크에서의 링크 예측에 초점을 맞추고 있다. 서명된 소셜 네트워크의 가장 중요한 특징은 그것이 포함하는 사회적 관계의 다양성이다.  서명된 소셜 네트워크에 관한 많은 기존 연구는 링크 예측 문제를 인접 매트릭스 완료 문제로 변환하는 사회 이론(예: 구조 균형 이론[16])에 기초하고 있다[16, 30].  • Q1: Which unconnected node pairs are likely to be connected  by a link in future?  • Q2: What will the signs of those new links be?  많은 서명된 소셜 네트워크가 사회 이론과 일치하기 때문에, 이러한 접근법은 새로운 링크의 징후(Q2)를 정확하게 예측하는 데 능숙하다.  그러나, 결과적인 예측은 비현실적인 가정에 근거한다. 즉, 우리는 어떤 노드가 연결될 것인지 이미 알고 있다.  따라서 사회이론에 기초한 접근방식은 일반적으로 어떤 두 개의 노드가 새로운 연결(Q1)을 형성할 것인지 예측할 수 없다.  서명된 소셜 네트워크와는 달리, 서명되지 않은 네트워크에서 새로운 링크를 예측하는 것은 본질적인 과제다 [1, 2, 4, 8, 31].  최근, 일부 학자들은 서명된 소셜 네트워크에서의 새로운 링크의 출현을 보다 충분히 고려하기 시작했다[18, 19, 26].  형상 기반 모델은 주로 사이클 구조[6]에 의존하며, 특히 3중 구조물에 의존한다.  촘촘한 네트워크에서는 사이클이 풍부해 효과적인 예측 모델을 비교적 쉽게 훈련할 수 있다.  그러나, 희박한 네트워크와 같이 사이클이 충분하지 않을 때는 성능이 저하된다.  불행히도, 대부분의 서명된 소셜 네트워크는 매우 희박하다. 따라서, 우리는 우리의 모델과 함께 작동하도록 두 가지 유형의 SLF를 설계했다. 하나는 양이고 다른 하나는 음이다. 각각은 사회적 관계의 형성에 대해 반대의 결과를 낳는다. 두 가지 유형의 SLF는 두 개의 독립적인 SLF 공간에 매핑되며, 여기서 각 치수는 SLF를 나타낸다. 또한 각 공간 내에서 각 노드는 각각 관심사와 구성을 나타내는 두 개의 SLF 벡터로 특징지어진다. 이러한 벡터는 관계의 방향성을 포착하는 데 사용된다. 전체적인 접근방식은 SLF 모델을 위해 특별히 설계된 3단계 절차로 캡슐화된다. 비선형 요소와 사회학적 의미는 노드 쌍에 대한 4가지 유형의 SLF 점수를 생성하는데 사용되며, 각각은 모델에서 고려된 4가지 유형의 사회적 관계에 해당된다. 이 네 가지 관계는 긍정적이고 부정적이며 중립적이며 전혀 관계가 없다. 긍정적인 관계와 부정적인 관계가 가장 일반적으로 고려된다. 그러나 중립적인 사회관계는 또한 한 노드가 다른 노드와 연결하려는 경향을 반영할 수 있다 [3, 24, 25]. 예를 들어, 선거 네트워크[5]에서, 우리는 일반적으로 중립적인 표를 많이 던진 적극적 유권자가 투표하지 않기로 선택하기보다는 앞으로도 계속 그러한 방식으로 투표할 것이라고 믿는다. 고려한다면, 이 정보는 예측 결과에 영향을 미칠 것이다. 그러나 대부분의 기존 접근방식은 중립적인 사회적 관계를 무시하거나 "무관계" 바구니로 통합한다. 일반적인 결과는 모형이 다른 노드와 연결하려는 노드의 성향을 과소평가하는 것이다. 이와는 대조적으로, 본 논문에서 제시된 SLF 모델은 중립적인 관계를 잘 활용한다. SLF 벡터는 음의 로그 우도 객관적 함수를 최소화하여 학습하는데, 이 함수는 양의 관계와 음의 관계뿐만 아니라 중립 관계와 부재 관계도 고려한다. 또한 중립적 연계에 배정된 점수는 합리적이고 사회학적 의미를 갖는다. 관계가 존재하지 않는 null 관계는 또 다른 중요한 고려사항이다. 특히 대부분의 소셜 네트워크, 심지어 밀도가 높은 관계도 많은 null 관계를 포함하고 있기 때문이다. null 관계를 최적화하는 것은 시간이 많이 걸리지만 이러한 관계들 중 많은 것들이 SLF 벡터를 배우는 과정에서 비정보적이다. 따라서 노드 쌍 샘플링 프로세스를 최적화 절차에 통합했다.  우리의 실험은 4개의 실제 서명된 소셜 네트워크에 연결 예측 작업을 포함한다. 그 결과는 우리가 제안한 SLF 모델이 최첨단 방법에 비해 우월함을 보여준다.  4개의 실제 서명된 소셜 네트워크가 이러한 실험에 사용되었다. WikiElec, WikiRfa, Slashdot and Epinions. 4개의 네트워크는 모두 SNAP 저장소 2에서 수집되며, 4개의 네트워크 모두 매우 희박하다. WikiElec 및 WikiRfa [5]: 위키엘렉트는 위키백과 관리자 선거의 투표 네트워크이고 WikiRfa는 WikiElec보다 최신 버전이다.  이 두 데이터 집합은 각각 평균 스피어스가 31.7과 25.3인 긍정, 부정, 중립 투표(링크)를 포함하고 있다.  슬래시닷(Slashdot)은 기술 관련 뉴스 웹사이트[16]의 친선 네트워크다. 평균 경사는 13.4이다. 각 사용자는 다른 사용자를 친구(긍정적) 또는 적(부정적)으로 태그할 수 있다.  에피니언스는 소비자 검토 사이트의 신뢰 네트워크[16]이다. 평균 경사는 12.8이다. 회원은 리뷰의 질에 따라 타 이용자를 신뢰(긍정)하거나 불신(부정)할 수 있다.  비교 방법은 Scalable embeddings for signed networks (SIGNet), Matrix factorization (MF), Link-oriented signed network embedding (LSNE), Signed directed network embedding (SIDE)이다. 자기비교로서 SLF 디그레이드라고 표기되는 중립 링크를 고려하지 않는 SLF 모델도 시험했다.  SLF는 SLF의 성능이 경쟁력 있는 마이크로 F1을 제외한 모든 지표에서 가장 정확한 방법이었다.  이 결과는 제안된 SLF 모델의 효능을 뒷받침한다.  • SLF는 MF보다 훨씬 나은 결과를 제공했다.  MF는 잠재 인자 벡터의 양/음수 값에 따라 잠재 인자가 사회 관계에 미치는 반대 효과를 나타내는 전통적인 잠재 인자 모델이다.  이러한 결과는 SLF 모델에 의해 학습된 서명된 잠재 인자 벡터가 링크 예측 작업을 통해 성능을 크게 개선할 수 있는 잠재력을 가지고 있음을 시사한다.  • SIGNet, LSNE 및 SIDE는 임베딩 공간에 구조 근사치를 보존하여 각 노드의 임베딩 벡터를 학습한다.  SIDE는 SLF의 성능에 접근하는 유일한 방법이었고 마이크로 F1 결과가 더 좋았다.  하지만 SIDE의 전체 성적은 SLF보다 좋지 않았다.  이러한 결과는 노드와 서명된 잠재 인자 사이의 관계를 나타내는 서명된 잠재 인자 벡터가 링크 예측 작업에 더 적합할 수 있음을 시사한다.  • SIGNet, MF, LSNE는 다른 네트워크보다 Epinion에서 더 우수한 성능을 보였다.  에피니언스는 긍정적인 링크의 비율이 가장 높으며, 이는 이러한 방법이 서명된 소셜 네트워크에서 부정적인 링크를 예측하는 데 어려움을 겪을 수 있음을 나타낸다.  • 세 가지 방법 모두 서명된 소셜 네트워크의 임베딩 학습을 기반으로 함에도 불구하고 LSNE는 SIGNet이나 SIDE만큼 경쟁력이 없었다.  그러나 LSNE는 1, 2차 대리점을 보존함으로써 임베딩을 배운다.  따라서 이러한 결과는 이것이 연결 예측에 적절한 기법이 아님을 시사한다.  • SLF는 네 가지 지표 모두에서 SLF-degrade를 능가하여 중립 관계를 고려할 때 링크 예측 성능이 향상된다는 우리의 추측을 뒷받침했다.  • 또한, AUC@non 결과는 중립적 관계에서 가장 많은 이익을 얻는데, 이는 중립적 관계를 고려하는 것이 다른 노드와 연결하려는 노드의 성향에 대한 예측을 개선한다는 것을 나타낸다. |
| 논문 의의 | 본 논문에서 제시된 서명된 잠재 인자(SLF) 모델은 새로운 링크 예측, 링크의 징후 예측, 모든 첨단의 네트워크에서 정확한 예측 등 이러한 모든 과제를 극복하기 위해 설계되었다. 따라서, 우리는 우리의 모델과 함께 작동하도록 두 가지 유형의 SLF를 설계했다. 하나는 양이고 다른 하나는 음이다. 우리가 제안한 SLF 모델이 최첨단 방법에 비해 우월했다. |
| 논문의 한계점 및 기타 | SLF의 자기비교를 통해 중립적인 관계를 고려하는 것이 성과에 도움이 된다는 것을 알게 됐다.  4개의 실제 서명된 소셜 네트워크에서 몇 가지 최첨단 방법을 사용한 추가 실험은 SLF 모델이 서명된 소셜 네트워크의 예측 문제를 연결하기 위해 하는 발전을 보여준다. |