

개체들의 영구적인 특성을 고려하는 시간 지식 그래프 임베딩

이재현¹, 이연창², 김상욱^{2*}

¹한양대학교 인공지능학과

²한양대학교 컴퓨터소프트웨어학과

{dldja4, lyc0324, wook}@hanyang.ac.kr

On Exploiting Permanent Properties of Entities in Temporal Knowledge Graph Embedding

JaeHyun Lee¹, Yeon-Chang Lee², Sang-Wook Kim^{2*}

¹Dept. of Artificial Intelligence, Hanyang University

²Dept. of Computer Science, Hanyang University

요 약

시간 지식 그래프 임베딩 방법들은 주어진 시간 지식 그래프에 존재하는 개체 및 관계를 저차원의 임베딩 벡터로 표현하는 것을 목표로 한다. 그러나, 기존 방법들은 개체들의 임베딩 벡터에 그들의 시간에 따라 변화하는 특성을 반영하는 데에만 집중함에 따라, 그들의 영구적인 특성을 무시한다는 한계를 갖는다. 본 논문에서, 우리는 실세계 데이터 집합들을 이용한 실험을 통해, 시간 지식 그래프 임베딩에서 개체들의 영구적인 특성을 고려하는 것이 중요하다는 점을 논의한다.

1. 서론

지식 그래프 (knowledge graph, 이하 KG) 는 실세계에서 발생하는 사실 정보를 그래프 구조로 나타낸 지식 베이스이다. KG 의 우수한 표현력은 실세계 응용 (예: 추천시스템, 자연어처리) 의 정확도 개선에 도움이 되므로, 이러한 응용의 기반 기술로 활용될 수 있는 KG 임베딩 연구가 활발히 수행되어 오고 있다 [1]. KG 임베딩의 목표는 주어진 KG 에 존재하는 개체 및 관계를 저차원의 임베딩 벡터로 표현하는 것이다.

최근에는, KG 에 존재하는 많은 사실 정보들이 영구적으로 유효하지 않다는 특성을 고려하기 위해, 다양한 시간 지식 그래프 임베딩 (temporal knowledge graph embedding, 이하 TGKE) 방법들이 제안되어 오고 있다 [1, 2, 3, 4]. 그러나, 기존 TKGE 방법들은 개체들의 임베딩 벡터에 그들의 시간에 따라 변화하는 특성을 반영하는 데에만 집중함에 따라, 그들의 영구적인 특성을 무시하게 된다. 예를 들어, ‘한국’ 과 ‘일본’ 간의 지리적 특성은 시간에 따라 변하지 않는다.

앞서 언급한 기존 TKGE 방법들의 한계를 검증하기 위해, 우리는 먼저 개체들의 영구적인 특성을 고려할 수 있는 간단한 그래프 합성곱 신경망 기반의 TKGE 방법을 설계한다. 그 후, 실세계 데이터 집합들을 이용한 실험을 통해, 우리의 TKGE 방법과 기존 TKGE 방법들의 정확도를 비교한다. 이를 통해, 우리는 개체들의 임베딩 벡터에 그들의 영구적인 특성을 반영하는 것의 중요성을 논의하고자 한다.

2. 기존 TKGE 방법

최신 TKGE 방법들은 크게 (1) 행렬 분해 기반 방법, (2) 거리 기반 방법, 그리고 (3) 그래프 합성곱 신경망 기반 방법으로 분류될 수 있다.

먼저, 행렬 분해 기반 방법 [2] 은 TKG 를 (‘소스 개체’, ‘관계’, ‘타겟 개체’, ‘시간’) 에 대한 4 차원 텐서로 간주하고, 해당 텐서에 행렬 분해 기술을 적용하여 개체들, 관계들, 그리고 시간들을 저차원의 임베딩 벡터로 표현한다. 다음으로, 거리 기반 방법 [3] 은 TKG 에 존재하는 각 시간과 관계를 임베딩 공간 내

*교신 저자

의 기하학적인 변환 (geometric transformation) 으로 활용한다. 각 사건 정보가 주어지면, 소스 개체의 임베딩 벡터에 시간과 관계에 해당하는 변환을 적용시켰을 때 타겟 개체의 임베딩 벡터와 가까워지도록 개체, 시간, 그리고 관계의 임베딩 벡터들을 학습한다. 마지막으로, 그래프 합성곱 신경망 기반 방법 [4] 은 특정 시간에 각 개체와 연결된 이웃 정보 (즉, 개체 및 관계) 를 기반으로, 개체와 관계의 임베딩 벡터들을 학습한다.

3. 개체들의 영구적인 특성 고려하는 TKGE 방법

본 논문에서, 우리는 TKG 내 개체들의 영구적인 특성을 고려할 수 있는 간단한 그래프 합성곱 신경망 기반의 TKGE 방법인 StaticGCN 을 설계한다.

$$e_t^{l+1} = e_t^l + \sum_{(e_s, r_i) \in \bar{\mathcal{N}}_t(G)} (e_s^l \cdot r_i), \quad (1)$$

여기서, $e_t, e_s \in \mathcal{E}$, $r_i \in \mathcal{R}$ 이며 \mathcal{E} 는 TKG 내 개체들의 집합, \mathcal{R} 은 TKG 내 관계들의 집합이다. $\bar{\mathcal{N}}_t(G)$ 는 개체 e_t 가 타겟 개체로서 등장한 사실 정보들의 집합이다. e^l 은 l 번째 층 (layer) 에서의 개체 e 의 임베딩이며, r_i 는 i 번째 관계 r_i 의 임베딩이다. 개념적으로, StaticGCN 은 특정 시간과 관계없이 각 개체와 연결된 모든 이웃 정보를 기반으로, 개체와 관계의 임베딩 벡터들을 학습한다. 이를 통해, StaticGCN 을 통해 얻은 개체들의 임베딩 벡터는 그들의 특정 시간에서 갖는 특성이 아닌 영구적인 특성을 보존할 수 있다.

4. 실험

실험 환경. 우리는 다음 3 가지 실세계 TKG 데이터 집합들을 사용하여 실험을 수행한다: ICEWS14, ICEWS05-15(Integrated Crisis Early Warning System), YAGO(Yet Another Great Ontology). <표 1>은 본 논문에서 사용한 데이터 집합들의 통계를 보여준다.

<표 1> 데이터 집합 통계

데이터 집합	ICEWS14	ICEWS05-15	YAGO
개체	7,128	10,488	10,552
관계	230	251	10
시간	365	4,017	3,275
사실 정보	90,730	461,329	20,509

실험 방법. 우리는 3 가지 최신 TKGE 방법들 (즉, TNT-ComplEx [2], ChronoR [3], T-GAP [4]) 과 StaticGCN 의 정확도를 비교하는 실험을 수행한다. 우리는 각 방법의 정확도를 간선 예측 시나리오를 통해 측정한다. 간선 예측의 목표는 각 TKGE 방법이 주어진 TKG 에서 일부 제거된 사실 정보를 얼마나 정확하게 예측할 수 있는지를 평가하는 것이다. 우리는 정확도 척도로 mean reciprocal rank (MRR) 를 이용한다.

구체적으로, 우리는 먼저 각 데이터 집합을 트레이닝 (90%) 집합과 테스트 (10%) 집합으로 분할한다. 그 후, 트레이닝 집합을 기반으로 각 TKGE 방법을 수행하여, 개체들과 관계들의 임베딩 벡터들을 얻는

다. 이러한 임베딩 벡터들을 이용하여, 테스트 집합에 존재하는 각 사실 정보의 (1) 소스 개체 혹은 (2) 타겟 개체를 맞추는 실험을 수행한다 [2, 3, 4].

실험 결과. <표 2>는 앞서 언급한 간선 예측 시나리오를 통해 측정한 TNT-ComplEx [2], ChronoR [3], T-GAP [4], 그리고 Static GCN 의 정확도를 보여준다. <표 2>를 통해, 우리는 StaticGCN 이 기존 TKGE 방법들보다 우수한 정확도를 제공한다는 것을 확인할 수 있다. 이러한 결과는 개체마다 시간에 불변하는 특성이 존재한다는 것을 의미한다. 따라서, 본 논문에서 우리는 TKGE 방법들이 개체들의 임베딩 벡터에 그들의 영구적인 특성을 반영할 필요가 있다고 결론짓는다.

<표 2> 최신 방법들과 StaticGCN 의 간선 예측 정확도

데이터 집합	TNT-ComplEx	ChronoR	T-GAP	Static GCN
ICEWS14	0.560	0.625	0.610	0.637
ICEWS05-15	0.600	0.684	0.670	0.691
YAGO	0.180	-	0.164	0.227

-: 소스 코드 공개되지 않아 결과를 얻을 수 없었음

5. 결론

KG 에 존재하는 사실 정보들의 시간적 유효성을 고려하기 위해, 최근 다양한 TKGE 방법들이 제안되어 왔다. 본 논문에서, 우리는 기존 TKGE 방법들이 개체들의 영구적인 특성을 보존하지 못한다는 한계를 지적했다. 우리는 실세계 데이터 집합들을 이용한 실험을 통해, 개체들의 영구적인 특성을 고려할 수 있는 간단한 StaticGCN 이 기존 TKGE 방법들보다 우수한 정확도를 보인다는 것을 확인했다. 이러한 실험 결과는 TKGE 방법들이 개체들의 시간에 따라 변화하는 특성 뿐만 아니라 영구적인 특성을 함께 고려해야 한다는 것을 보여준다.

사사

이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단과 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. NRF-2020R1A2B5B03001960, No.2018R1A5A7059549, No.RS-2022-00155586).

참고문헌

- [1] S.Ji, S.Pan et al., "A Survey on Knowledge Graphs: Representation, Acquisition, and Applications," In *IEEE Transactions on Neural Networks and Learning Systems*, vol.33, no.2, pp. 494-514, 2022.
- [2] T.Lacroix et al., "Tensor Decompositions for Temporal Knowledge Base Completion," In *Proc. of the Int'l Conf. on Learning Representations*, 2020.
- [3] A.Sadeghian et al., "ChronoR: Rotation Based Temporal Knowledge Graph Embedding," In *Proc. of the Conf. on Artificial Intelligence*, 2021.
- [4] J.Jung et al., "Learning to Walk across Time for Temporal Knowledge Graph Completion," In *Proc. of the ACM Conf. on Knowledge Discovery & Data Mining*, 2021.