

개체 유형 정보를 활용한 다차원 지식 그래프 임베딩

박준혁*, 김장원**

Multi-dimensional Knowledge Graph Embedding with Entity Type Information

Junhyuk Park*, Jangwon Gim**

본 논문은 KISTI ScienceON 서비스의 학술정보 및 R&D 협업관리 기능(MyON)을 이용하여 작성되었습니다.

요 약

지식 그래프는 복잡한 지식 구조를 다루는 표현 방법으로 정보 검색 및 추천 시스템 등 다양한 분야에서 활용되고 있다. 지식 그래프 구축 과정 시 원천데이터에 포함된 개체들의 정보 및 개체 간의 관계를 정확하고 손실 없이 반영해야 한다. 본 논문에서는 개체 간의 관계 예측 성능 향상을 위하여 개체 유형 정보와 다차원 공간을 활용한 지식 그래프 임베딩 모델을 제안한다. 실험 결과 지식 그래프 임베딩 시 개체 유형에 따른 군집화 및 다차원 정보 표현 방식을 고려한 결과 베이스라인 모델보다 관계 예측의 성능이 최대 15%p 향상되어 제안 모델의 우수성을 보였다. 그러므로 개체 유형을 반영한 다차원 임베딩 모델은 향후 풍부한 지식 그래프 구축 및 다양한 관계를 포함하는 응용 분야의 지식 그래프 구축에 도움을 줄 것으로 기대한다.

Abstract

A knowledge graph is a representation method for handling complex knowledge structures and is used in various fields, such as information retrieval and recommendation systems. Therefore, in constructing such knowledge graphs, it is essential to accurately and without loss reflect the information of entities included in the source data and the relationships between entities. This paper proposes a knowledge graph embedding model that leverages entity-type information and multidimensional space to enhance relationship prediction performance between entities. Experimental results demonstrate up to a 15% improvement over baseline models, making this approach promising for constructing rich knowledge graphs with diverse relationships in various application domains.

Keywords

knowledge graph embedding, ChatGPT, entity type, negative sampling, non-euclidean geometry, link prediction

* 군산대학교 소프트웨어학부 학사과정
- ORCID: <https://orcid.org/0000-0003-1709-5678>
** 군산대학교 소프트웨어학부 교수(교신저자)
- ORCID: <https://orcid.org/0000-0002-4480-7944>

• Received: Oct. 27, 2023, Revised: Nov. 10, 2023, Accepted: Nov. 13, 2023
• Corresponding Author: Jangwon Gim
School of Software, Kunsan National University, 558, Daehak-ro, Gunsan.
Jeollabuk-do, Korea
Tel.: +82-63-469-8916, Email: jwgim@kunsan.ac.kr

1. 서론

지식 그래프는 다양한 개체 간의 관계를 표현하여 현실 세계의 복잡한 상호작용에 대한 모델링이 가능하다[1]. 이에 따라 최근 빅데이터 환경에서 다양한 도메인에서 정보의 체계적인 관리 및 데이터 간 의미적 통합을 위한 지식 모델링, 지식베이스 구축 및 도메인 지식 연결 등을 위해 지식 그래프 연구가 활발히 진행되고 있다[2]. 그렇지만 현실 세계의 정보를 그래프 구조로 표현하는 과정에서 개체의 오식별, 개체 간의 관계 추출 누락은 지식 그래프에 포함된 정보의 손실 및 불충분한 형태의 지식 그래프 생성을 야기한다[3]. 따라서 지식베이스의 역할로 데이터 연계 및 통합, 응용 서비스 개발 등에 적용하기 위한 지식 그래프는 의미적 불완전성으로 인한 문제를 발생시킬 수 있다. 최근 지식 그래프 임베딩은 지식 그래프에 포함된 개체 및 관계 정보를 벡터 공간에 사상하여 개체 간의 관계를 식별하고, 개체들이 가질 수 있는 개체 관계 정보를 예측하는데 적용되고 있다[4]. 이를 위해서 지식 그래프 임베딩에서는 복잡한 지식 그래프의 구조와 개체들의 상호 관계 정보를 벡터 공간으로 사상하여 벡터 간의 연산을 통해 개체 간의 유사성과 관계 추론을 가능하게 한다. 따라서 개체 간의 관계, 개체의 영향력, 관계의 가중치 등을 도출할 수 있으며 초거대 언어모델의 학습 성능 향상, 질의응답 기반의 정보 검색 및 추천 시스템 등의 응용 분야에서 활발히 활용되고 있다[5].

지식 그래프 임베딩의 성능을 높이고자 다양한 연구가 제안되었다. 그중 다차원 임베딩 기법은 임베딩 공간을 단일 유클리드 공간에서 비유클리드 공간으로 확장하여 개체 사이의 기하학적 상호작용 정보를 지식 그래프 임베딩의 학습에 반영한다[6][7]. 또한, 다차원 임베딩 기법을 계층적 데이터에 적용하여 SOTA(State-Of-The-Art) 성능을 보인 연구가 제안되었다[8].

II. 관련 연구

2.1 ChatGPT를 통한 개체 유형 추출

ChatGPT란 GPT-3.5 및 GPT-4 모델을 기반으로 진화하고 있는 대화형 인공지능 서비스이다. GPT-4는 OpenAI가 개발한 GPT 시리즈의 4번째 언어모델로 텍스트와 시각적 입력을 처리하여 답변을 생성할 수 있는 인공지능 언어 모델이다[9]. 최근 GPT-4 버전이 적용된 ChatGPT는 기존 챗봇 시스템의 성능을 뛰어넘어 인간과 자연스러운 대화를 지원할 수 있는 정도로 향상된 성능을 보인다[10]. 이러한 ChatGPT는 단일 프롬프트 입력 및 연속적인 프롬프트의 입력에 따라 제공하는 응답을 다르게 생성한다. 따라서 응답 결과의 만족도 향상을 위해 CoT(Chain of Thought)와 같은 연속적인 프롬프트의 입력을 통해 사용자 질문에 대한 응답 성능 향상을 위한 연구가 활발히 진행되고 있다[11]. 최근 방대한 양의 데이터를 학습한 ChatGPT를 대상으로 한 프롬프트에는 일반적인 자연어 질문에 응답을 기대하는 것뿐만 아니라, 다양한 프로그래밍 언어에 코드 생성, 파라미터 입력을 통한 자동 이미지 생성 지원 등에서 적용되고 있다. 또한, 지식 그래프에 포함된 개체 정보를 프롬프트에 입력하여 ChatGPT로부터 입력된 개체의 유형 정보를 획득할 수 있다.

2.2 비유클리드 공간 임베딩 모델

비유클리드 공간은 유클리드 기하학이 갖는 복잡한 기하학적 특성들을 나타내지 않는 공간을 의미한다[12]. 쌍곡면(Hyperbolic) 기하학은 평행선이 무한대로 향하는 특성을 갖추며, 평면 위의 삼각형에서 내각의 합이 180도보다 작아지는 특이한 특성을 보인다. 이는 쌍곡면 공간이 유클리드 공간보다 더 큰 용량을 가진다는 것을 의미한다. MuRP는 쌍곡면 공간의 다중 관계 그래프를 포함하는 모델을 제안함으로써 유클리드 공간 임베딩 모델과 관계 예측 실험에서 우수한 결과를 보였다[7]. 그렇지만 지식 그래프의 관계가 계층적이지 않을수록 성능이 저하되는 모습을 보인다. 초구(Hyper-spherical) 공간은 직선이 무한히 확장되는 유클리드 기하학과 달리 초구체의 표면을 따라 이동하는 경로를 고려한 공간이다[13]. 유클리드 기하학에서 평행선은 무한히 확장되어도 서로 교차하지 않지만, 초구 공간에서 평행선은 곡률이 존재하기 때문에 평행선이 서

서히 가까워지는 성질이 있다. 이러한 초구 공간의 성질은 모든 점이 초구의 중심에서 동일 거리에 있다는 특징을 보이며, 이러한 초구 공간의 특징은 데이터들의 관계 예측에 효과적인 ManifoldE 모델을 제안하였다[6]. ManifoldE는 개체와 관계를 초구면 공간과 같은 연속 벡터 공간으로 변환하여 임베딩을 생성하여 유클리드 공간 임베딩 모델보다 관계 예측 성능이 향상됨을 보인다. 그렇지만 벡터 공간별 기하학적 정보가 서로 영향을 받지 않도록 각 공간별로 사상된 벡터 정보만을 사용하고 있다[6]. 따라서 다양한 공간에 사상될 수 있는 지식 그래프를 보다 정확히 표현하기 위하여 개별 공간만을 이용하는 것이 아닌 초구 공간의 특징, 즉 다차원 공간을 반영한 지식 그래프 임베딩 기법이 필요하다.

2.3 네거티브 샘플링

네거티브 샘플링은 머신 러닝에서 확률적인 샘플링 기법 중 하나로, 주로 대량의 데이터셋에서 효과적인 학습을 위해 사용된다[14]. 지식 그래프 임베딩 모델의 학습 시 네거티브 샘플링은 주어, 술어, 목적어 구조로 이루어진 트리플 데이터에서 주어 및 목적어에 대응하는 개체를 오답 개체로 대체하여 오답 트리플 데이터를 생성하는 기법이다. 즉, 개체 간 정보의 손실 및 오답을 포함한 지식 그래프에서 네거티브 샘플링 기법을 적용하여 지식 그

래프에 포함된 개체 간 관계 예측 성능 향상에 유의한 효과를 보인다. 따라서 계층 구조를 포함한 데이터는 지식 그래프 내에서 추론 작업을 수행할 때 상위 개념과 하위 개념 간의 관계를 고려하여 추론 정확도를 향상시킬 수 있다.

III. 제안 모델

관계 예측의 성능 향상을 위해서 개체의 유형 정보와 다차원 기법을 지식 그래프 임베딩에 적용한 모델을 제안한다. 그림 1은 제안 모델의 전체 흐름도를 나타낸다.

3.1 개체 유형 추출

지식 그래프에 포함된 개체 유형 식별을 위해 ChatGPT를 이용한다. 각각의 개체들은 개체 유형을 가질 수 있으며 개체들을 개체 유형별로 그룹화 할 수 있다. 이러한 개체 유형정보를 적용하여 지식 그래프 임베딩에 적용하여 관계 예측 성능을 향상한 모델이 제안되었다[15]. 개별 개체에 대한 유형 식별을 위해 제안 모델에서는 개체명과 개체 정의(문자열)를 함께 ChatGPT의 프롬프트에 입력한다. 이때 개체 정의는 동음이의어와 같은 개체명 식별을 위해 사용한다. 그림 2는 ChatGPT를 통해 각각의 개체에 대한 개체 유형 추출 예를 보인다.

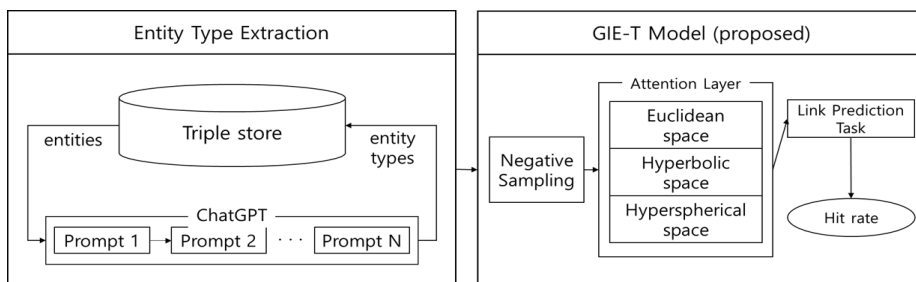


그림 1. 제안 모델 개요도

Fig. 1. Overview of the proposed model

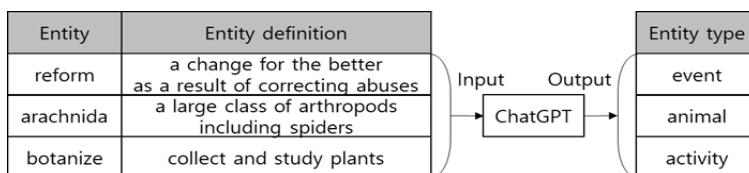


그림 2. ChatGPT를 활용한 개체 유형 추출 예

Fig. 2. Example of extracting entity types using ChatGPT

3.2 네거티브 샘플링

지식 그래프 임베딩의 성능 향상을 위해 제안 모델은 네거티브 샘플링 기법을 적용한다. 네거티브 샘플링을 통한 오답 데이터 생성 시, 무작위로 개체를 추출하여 오답 트리플 데이터를 생성하므로 생성된 트리플은 개체의 유형을 포함하지 않은, 즉 트리플을 구성하는 개체 간의 유형 정보가 반영되지 않는다. 그림 3은 개체 유형 정보를 고려한 네거티브 샘플링의 예를 보인다.

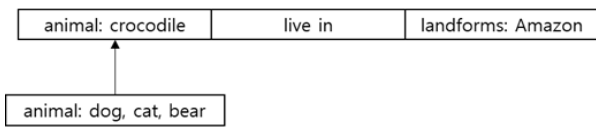


그림 3. 네거티브 샘플링 예시
Fig. 3. Example of negative sampling

예를 들어 한 개의 트리플이 주어, 술어, 목적어 형태로 각각 주어('animal: crocodile'), 술어('live in'), 목적어('landforms: amazon')이며 이에 대한 오답 데이터 생성 시 'crocodile'의 특징을 나타내는 개체 유형 정보('animal')를 참조하면 동일 유형('animal')에 포함되는 'dog' 개체로 오답을 생성한다.

따라서 제안 모델에서는 이러한 개체별 유형 정보를 지식 그래프 임베딩에 반영하기 위한 손실함수를 식 (1)과 같이 정의한다. 집합 T 는 개체 유형 t 를 원소로 포함하며, V_t 는 개체 유형 t 에 속하는 인스턴스 집합을 의미하며 e 는 개체 유형 t 에 속하는 개체의 임베딩 값을 의미한다. 이때 e_i 는 한 개의 'dog' 개체의 임베딩 값을 의미하며 'Animal'에 속한 개체들의 임베딩 평균값은 avg_t 로 정의한다. 그러므로 네거티브 샘플링 시 동일 유형 내에 포함된 임의의 개체(e_j)를 추출하기 위해서 L_t 의 값을 0으로 수렴하도록 한다.

$$L_t = \sum_{i \in T} \sum_{e \in V_t} \|e_i - avg_t\| \quad (1)$$

3.3 다차원 임베딩

지식 그래프 임베딩 시 벡터 공간의 기하학적 특

정에 따라 사상되는 벡터 정보의 파악이 중요하다. [8]은 이러한 특징을 기반으로 다차원 임베딩의 성능이 단일 공간의 사상보다 우수함을 보인다. 또한, 다차원의 서로 다른 공간의 특징을 추출하여 상호 작용하는 기하학적 사상 정보에 어텐션 메커니즘을 적용한 임베딩 기법이 제안되었다[8]. 개체 정보에 대한 지식 그래프 임베딩 시, 기하학적 특징을 반영한 각 차원의 임베딩 값을 손실함수에 반영하기 위해 식 (2), (3)을 정의한다. E, H, S 는 각각 유클리드, 쌍곡면, 초구면 공간의 임베딩 값을 의미하며 h, r, t 는 주어, 관계, 목적어 개체를 의미한다. 이때 λ, α 는 어텐션 가중치 값을 나타낸다. 따라서 식 (2), (3)은 주어, 목적어 개체의 서로 다른 공간 임베딩 값에 가중치를 부여하여 상호작용하는 기하학적 정보를 계산한다.

$$Hdist(E_h, H_h, S_h) = \exp(\lambda_E E_h + \lambda_H \log(H_h) + \lambda_S \log(S_h)) \quad (2)$$

$$Hdist(E_t, H_t, S_t) = \exp(\alpha_E E_t + \alpha_H \log(H_t) + \alpha_S \log(S_t)) \quad (3)$$

3.4 손실 함수

제안 모델에서는 개체 유형 정보를 반영한 네거티브 샘플링과 다차원 임베딩 기법을 적용하기 위한 손실 함수를 정의한다. 식 (4)는 기하학적 다형성과 개체의 유형 정보를 임베딩에 반영하기 위한 수식이다.

$$\Phi(h, r, t) = -(d_c(Hdist(E_h, H_h, S_h), t) + d_c(Hdist(E_t, H_t, S_t), h)) + b \quad (4)$$

식 (4)의 d_c 는 기하학적 공간에 사상된 주어(h)와 목적어(t) 개체의 유클리디안 거리를 계산하는 수식이며 b 는 편향을 의미한다. 그러므로 식 (2)을 통한 주어의 임베딩과 목적어의 거리, 식 (3)을 통한 목적어의 임베딩과 주어의 거리를 계산한다.

$$L_g = \sum \log(1 + \exp(-Y\Phi(h_i, r_j, t_k))) \quad (5)$$

식 (5)에서 Y 는 트리플을 구성하는 h_i, r_j, t_k 가 정답일 경우 +1, 오답이면 -1을 가진다. 이때 트리플은 r_j 를 대상으로 h_i 와 t_k 개체가 달라지며 r_j 에 대한 정답(h_i, t_k)을 찾기 위해 양의 방향으로 학습한다. 따라서 개체 유형을 반영한 네거티브 샘플링 모델과 기하학적 특징을 반영한 다차원 임베딩 모델의 가중치를 고려한 손실함수를 식 (6)과 같이 정의한다. α 는 개체 유형 정보(L_t)의 β 는 다차원 공간 정보(L_g)에 대한 각각의 가중치이다.

$$L = \alpha L_t + \beta L_g \quad (6)$$

IV. 실험

4.1 실험 환경

표 1은 제안 모델의 실험에 사용된 서버 사양이다. 식 (6)에 정의된 α, β 의 최적값 도출을 위한 실험 결과 최대 성능은 각각 0.3과 0.7로 나타났다.

표 1. 실험 환경

Table 1. Experiment environment

OS	Ubuntu 18.04.06 LTS
CPU	Intel(R) Xeon(R) Silver 4210R CPU 2.40GHz
Memory	64GB
GPU	RTX A6000 (48GB, x 2)

4.2 실험 데이터

WN18RR은 대규모 영어 어휘 데이터베이스를 기반으로 하는 WordNet의 하위 집합으로 지식 그래프의 관계 예측 실험에서 대표적으로 사용되는 데이터셋이다. WN18RR 데이터셋은 40,943개의 개체와 11개의 개체들의 관계로 구성된 총 93,003개의 트리플 데이터셋이다.

WN18RR에 포함된 개체를 대상으로 ChatGPT를 이용한 개체 유형 식별 결과 총 5,677개의 개체 유형이 도출되었다. 따라서, 추출한 개체 유형에 포함된 개체 및 그들의 관계를 기존 WN18RR에서 추출한 결과 38,774개의 개체와 11개의 관계 어휘가 포

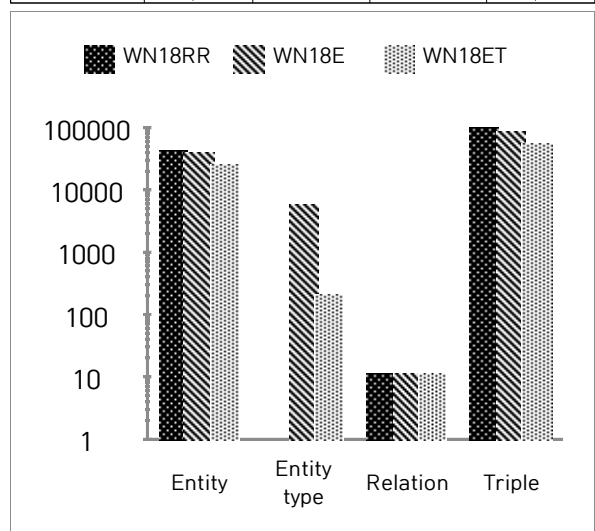
함된 총 82,228개의 트리플 데이터를 구축하여 WN18E로 정의하였다. 이때, WN18E 트리플 데이터의 개체 유형 정확도 검증을 위해 개체 유형별로 개체 5,000건을 랜덤하게 추출하였으며 레이블링의 경험을 보유한 AI 연구자(3명)에 의한 교차 검증을 실시하였다. 그 결과 각각 97.58%, 97.50%, 98.50% (평균 97.86%)의 정확도를 보여서 ChatGPT를 적용한 개체 유형 식별이 유의함을 확인하였다.

또한, WN18ET는 대표적인 개체 유형에 포함된 개체들의 개수에 따른 영향력을 검증하기 위해 구축한 데이터셋으로서 유형별 개체의 고빈도 및 저빈도를 고려하였다. 이를 위해 임계치(70%)를 지정하여 전체 개체 건수 대비 고빈도 개체를 포함하고 있는 유형을 선정한 데이터셋이다. 따라서 WN18ET 데이터셋은 WN18E 데이터셋의 서브셋으로 24,012개의 개체와 202개의 개체 유형, 11개의 관계를 포함하고 총 52,063개의 트리플 데이터를 가진다. 표 2는 WN18RR, WN18E, WN18ET에 대한 각 개체, 관계, 트리플 건수를 나타내며 로그 스케일을 적용한 그래프를 함께 표기하였다.

표 2. WN18RR, WN18E, WN18ET의 개체, 개체 유형, 관계, 트리플 수

Table 2. Number of entities, entity types, relations, total triples in WN18RR, WN18E, WN18ET

Info. Datasets	Entity	Entity type	Relation	Triple
WN18RR	40,943	-	11	93,003
WN18E	38,774	5,677	11	82,228
WN18ET	24,012	202	11	52,063



4.3 실험 평가

표 3은 기존 연구인 TransE_type[16], GIE[8] 및 제안 모델(GIE-T)에 대한 관계 예측 실험 결과를 보인다. 각 모델에 대한 비교 평가를 위해 적중률(Hit Rate@K)을 사용한다. 적중률은 K개의 예측 값 중에서 실제 값과 일치하는 비율을 나타낸다. 예를 들어 k=3인 hit3은 3개의 예측값 중에서 실제값이 나타난 비율을 의미한다.

표 3. TransE_type, GIE, 제안 모델의 비교 평가 결과
Table 3. Comparison results among TransE_type, GIE and proposed model

Hit rate \ Model	TransE_type [16]		GIE[8]		GIE-T (proposed)	
	WN 18E	WN 18ET	WN 18E	WN 18ET	WN 18E	WN 18ET
hit1	0.285	0.079	0.381	0.083	0.359	0.146
hit3	0.326	0.166	0.504	0.197	0.507	0.217
hit10	0.494	0.213	0.574	0.265	0.643	0.381

제안 모델은 hit10 대상으로 한 결과에서 TransE_type, GIE보다 각각 약 15%p, 7%p 우수한 성능을 보였다. 또한 70% 수준의 임계치가 적용된 WN18ET 데이터는 기존 모델보다 각각 약 17%p, 12%p 높은 성능을 보였다. 따라서 개체 유형과 다차원 임베딩 기법을 적용한 제안 모델의 관계 예측 성능이 우수함을 확인할 수 있다.

V. 결론 및 향후 연구

관계 예측은 지식 그래프의 활용 및 서비스 분야에서 부족한 데이터의 확장 및 추론 등을 위해 중요한 역할을 한다. 제안 모델은 개체 유형과 다차원 임베딩 기법을 적용하여 지식 그래프에 포함된 개체의 관계 예측 성능이 기존 모델보다 향상됨을 보였다. 또한, ChatGPT 프롬프트를 이용한 개체명 유형 식별 및 정제를 통해 개체 유형 정보 생성이 가능함을 보였다. 그 결과, 제안 모델은 군집화 특성을 가진 데이터들을 대상으로 제안 모델의 적용을 통해 유의한 결과를 도출할 수 있을 것으로 기대할 수 있다. 그러므로 추천 시스템 및 유사한 감정 유형에

대한 분류에서 제안 모델을 적용할 수 있을 것으로 기대한다. 따라서 초거대 모델을 적용한 개체 유형 정답 집합 생성 자동화, 대용량의 응용 도메인 지식 그래프에 대한 관계 예측을 향후 연구로 한다.

References

- [1] D. Fensel, U. Simsek, K. Angele, E. Huaman, E. Karle, O. Panasiuk, I. Toma, J. Umbrich, and A. Wahler, "Introduction: what is a knowledge graph?", Knowledge graphs: Methodology, tools and selected use cases, pp. 1-10, Feb. 2020. https://doi.org/10.1007/978-3-030-37439-6_1.
- [2] M. Kejriwal and P. Szekely, "Knowledge graphs for social good: an entity-centric search engine for the human trafficking domain", IEEE Transactions on Big Data, Vol. 8, No. 3, pp. 592-606, Jun. 2022. <https://doi.org/10.1109/TBDDATA.2017.2763164>.
- [3] S. J. Choi and S. Park, "A Knowledge Graph Embedding-based Ensemble Model for Link Prediction", Journal of KIISE, Vol. 47, No. 5, pp. 473-478, May 2020. <https://doi.org/10.5626/JOK.2020.47.5.473>.
- [4] M. Hwang and J. J. Whang, "Knowledge Graph Embedding with Dynamic Attention", Proceedings of Korea Software Congress 2022 (KSC 2022), pp. 1648-1650, Jun. 2022.
- [5] S. Wang, et al., "Mixed-curvature multi-relational graph neural network for knowledge graph completion", Proceedings of the Web Conference 2021, New York, USA, pp. 1761-1771, Apr. 2021. <https://doi.org/10.1145/3442381.3450118>.
- [6] H. Xiao, M. Huang, and X. Zhu, "From one point to a manifold: knowledge graph embedding for precise link prediction", Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence (IJCAI'16), New York, USA, pp. 1315-1321, Jul. 2016.
- [7] I. Balazevic, C. Allen, and T. Hospedales, "Multi-relational poincare graph embeddings",

- Advances in Neural Information Processing Systems, Vol. 32, pp. 4463-4473, Oct. 2019. <https://doi.org/10.48550/arXiv.1905.09791>.
- [8] Z. Cao, Q. Xu, Z. Yang, X. Cao, and Q. Huang, "Geometry interaction knowledge graph embeddings", Proceedings of the AAAI Conference on Artificial Intelligence, California, USA, Vol. 36, No. 5, Jun. 2022. <https://doi.org/10.48550/arXiv.2206.12418>.
- [9] OpenAI, <https://openai.com/gpt-4> [accessed: Oct. 27, 2023]
- [10] H. Liu, R. Ning, Z. Teng, J. Liu, Q. Zhou, and Y. Zhang, "Evaluating the logical reasoning ability of ChatGPT and GPT-4", arXiv preprint arXiv:2304.03439, Apr. 2023. <https://doi.org/10.48550/arXiv.2304.03439>.
- [11] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. H. Chi, Q. V. Le, and D. Zhou, "Chain-of-thought prompting elicits reasoning in large language models", Advances in Neural Information Processing Systems, Vol. 35, pp. 24824-24837, Dec. 2022.
- [12] N. A. Asif, et al., "Graph neural network: A comprehensive review on non-euclidean space", IEEE Access, Vol. 9, pp. 60588-60606, Apr. 2021. <https://doi.org/10.1109/ACCESS.2021.3071274>.
- [13] R. C Wilson, E. R Hancock, E. Pekalska, and R. P.W. Duin, "Spherical and hyperbolic embeddings of data", IEEE transactions on pattern analysis and machine intelligence, Vol. 36, No. 11, pp. 2255-2269, Apr. 2014. <https://doi.org/10.1109/TPAMI.2014.2316836>.
- [14] Z. Yang, M. Ding, C. Zhou, H. Yang, J. Zhou, and J. Tang, "Understanding negative sampling in graph representation learning", Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining, pp. 1666-1676, New York, USA, Aug. 2020. <https://doi.org/10.1145/3394486.3403218>.
- [15] R. Xie, Z. Liu, and M. Sun, "Representation learning of knowledge graphs with hierarchical types", Proceedings of the 25th International Joint Conference on Artificial Intelligence, New York, USA, Vol. 2016, pp. 2965-2971, Jul. 2016.
- [16] S. Kong, C. Chung, S. Ju, and J. J. Whang, "Knowledge Graph Embedding with Entity Type Constraints", Journal of KIISE, Vol. 49, No. 9, pp. 773-779, Sep. 2022. <https://doi.org/10.5626/JOK.2022.49.9.773>.

저자소개

박 준 혁 (Junhyuk Park)



2017년 3월 ~ 현재 : 군산대학교
소프트웨어학부 학사과정
관심분야 : 자연어처리, 지식그래프
임베딩 및 빅데이터 분석

김 장 원 (Jangwon Gim)



2008년 2월 : 고려대학교
컴퓨터학과(이학석사)
2012년 8월 : 고려대학교
컴퓨터·전파·통신공학과(공학박사)
2013년 3월 ~ 2017년 3월 :
한국과학기술정보연구원
선임연구원

2017년 4월 ~ 현재 : 국립군산대학교 소프트웨어학부
부교수

관심분야 : 자연어처리, 지식그래프, 지식그래프 임베딩,
실시간 빅데이터 분석