
지식 그래프를 이용한 뉴럴 엔터티 링크

2020.12.20

전북대학교 홍승연

목차

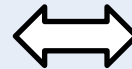
- Background
- Related Works
- Model
- Experiment Result
- Conclusion

개체명 연결 (Entity Linking)

1. Local Model

- Input pair(Context, Mention)과 Candidate Entities의 유사도 비교

Context Mention



Entities

Compare
Similarities

가수 이승철은 존박의 1집 앨범 '이너차일드(Inner Child)'에 대해 ...

Candidate Entities

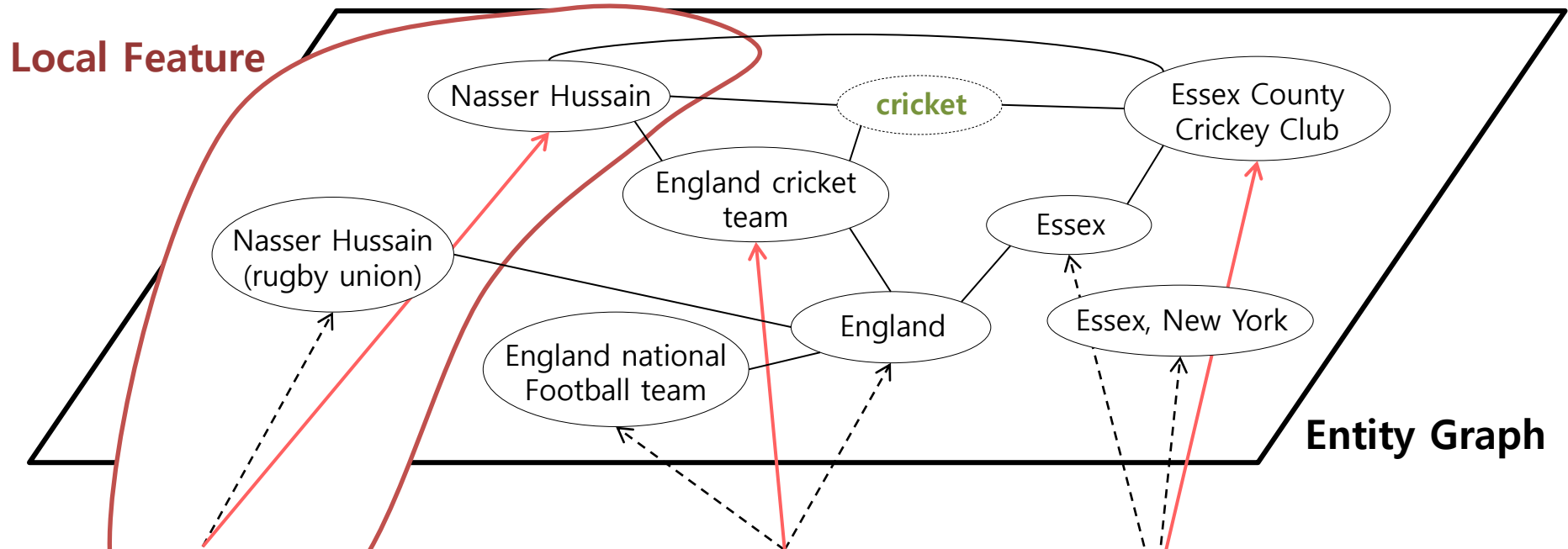
이승철_(가수)	0.65
이승철_(배우)	0.05
이승철_(기업인)	0.01
...	

2. Global Model

모든 Mention의 Entity가 통일성(coherence) 있게 연결 되도록 하는 방법

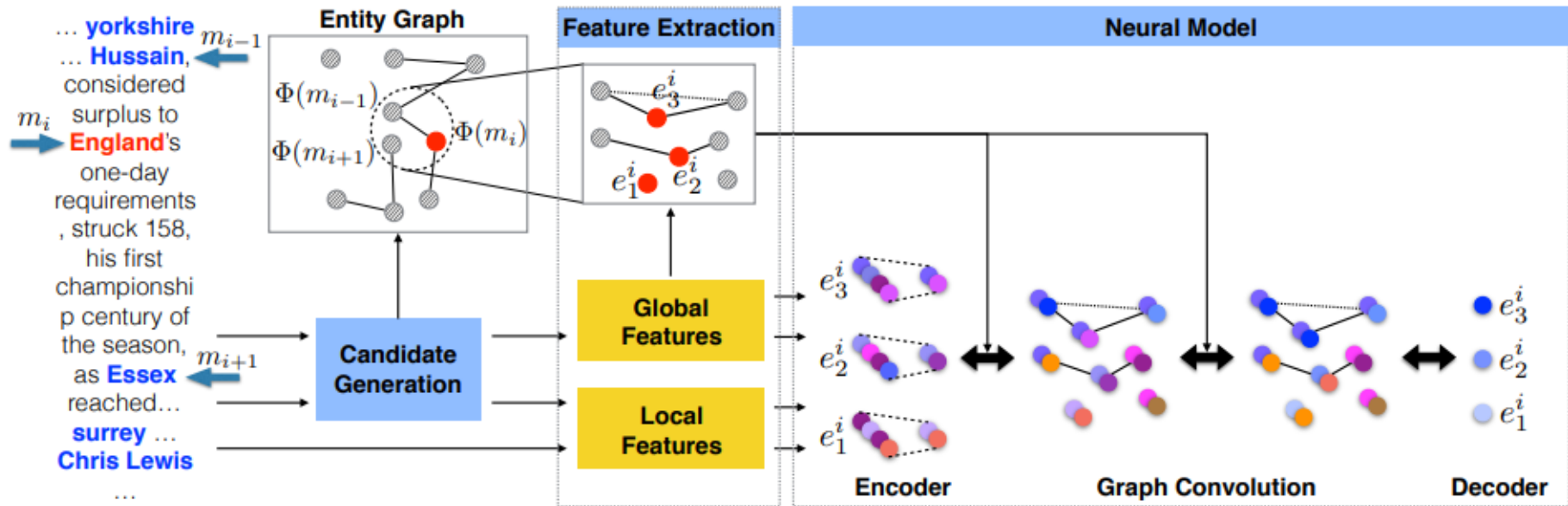
Neural Collective Entity Linking [Y. Cao, 2018]

Local Feature + Global Feature



Hussain, considered surplus to **England**'s one-day requirements, Struck 158, his first championship century of the season, as **Essex** reached 372 And took a first innings lead of 82

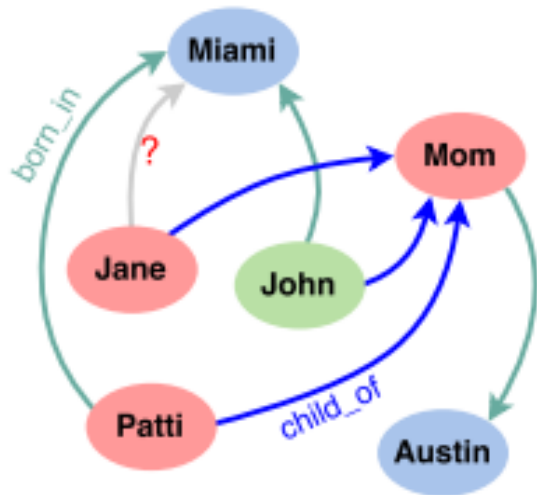
Neural Collective Entity Linking [Y. Cao, 2018]



- 그래프 구조에 문맥 정보가 반영되지 않는 문제가 존재(그래프 구조 고정)
- 개체간의 관계가 제대로 학습 이루어졌다고 하기 어려움

A survey of embedding models of entities and relationships for knowledge graph completion[Nguyen' 2020]

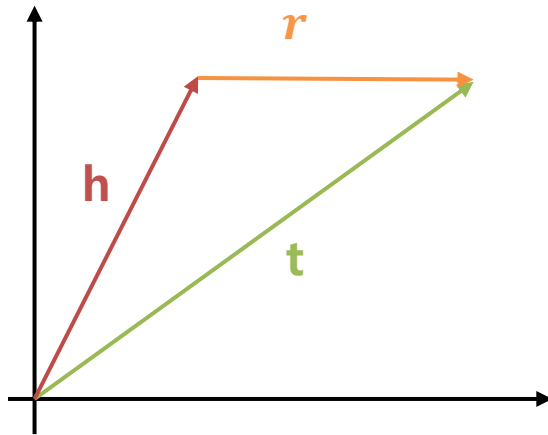
지식 그래프는 트리플 구조로 표현된 real-world의 지식 집합



$f(h, r, t) > f(h', r', t')$
 $f(h, r, t)$ correct Triple
 $f(h', r', t')$ incorrect Triple

- 지식 그래프 임베딩에 기반한 학습은 지식 베이스가 불완전하기 때문에 지식 그래프 완성 문제를 풀기 위해 제안
- 임베딩 기반 방법은 Positive Triple Score가 Negative Triple Score보다 커지도록 학습하는 것을 목표로함
- f 함수를 통해 트리플 점수를 구하고 다양하게 정의 가능

Translating Embeddings for Modeling Multi-relational Data [Bordes' 2020]



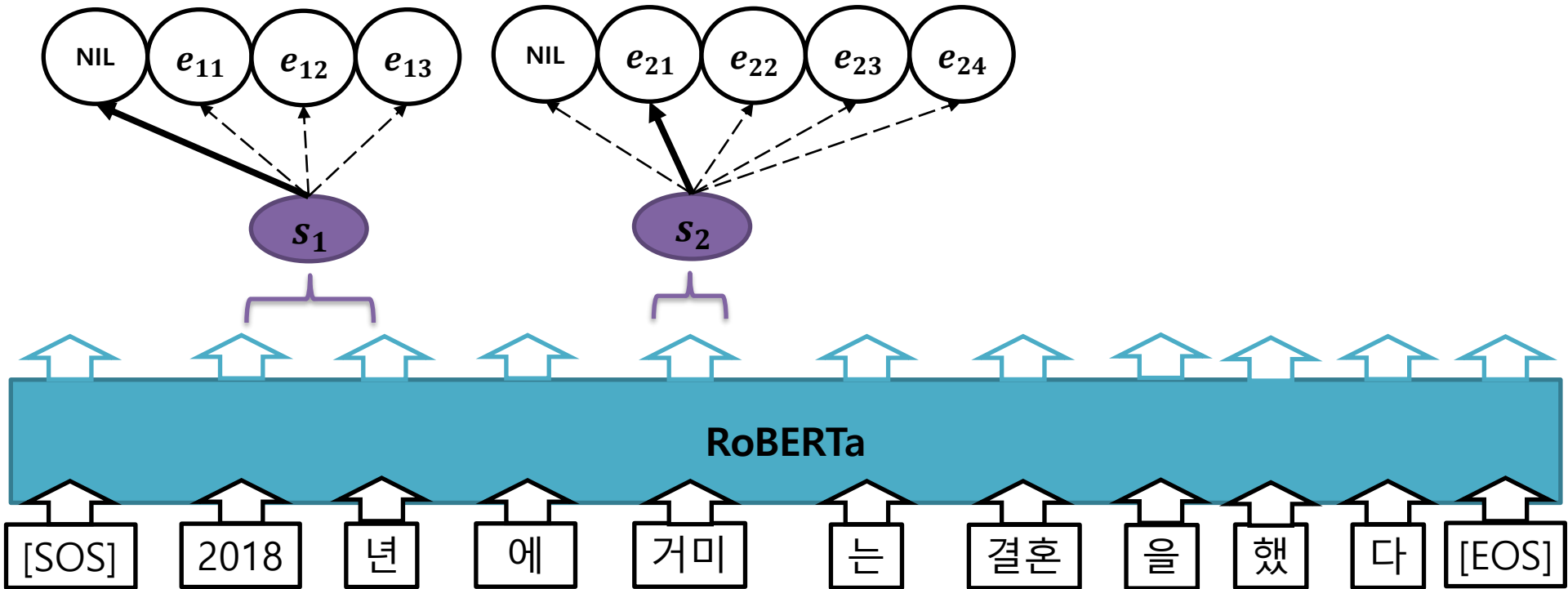
$$\begin{aligned} & - \|v_{\text{서울}} + v_{\text{수도}} - v_{\text{대한민국}}\| > \\ & - \|v_{\text{서울}} + v_{\text{수도}} - v_{\text{미국}}\| \end{aligned}$$

$$\mathcal{L} = \sum_{\substack{(h,r,t) \in S \\ (h',r,t') \in S'}} [Y - d(h,r,t) + d(h',r,t')]_+$$

$$[x]_+ = \max(0, x)$$

- 벡터간의 연산 관점에서 벡터 거리를 통한 score함수 사용
- f 함수로 벡터 거리를 구하는 연산 사용(L1 norm, L2 norm)

멘션 임베딩을 이용한 NIL 멘션 탐지와 개체 연결의 통합 모델[홍승연'20]

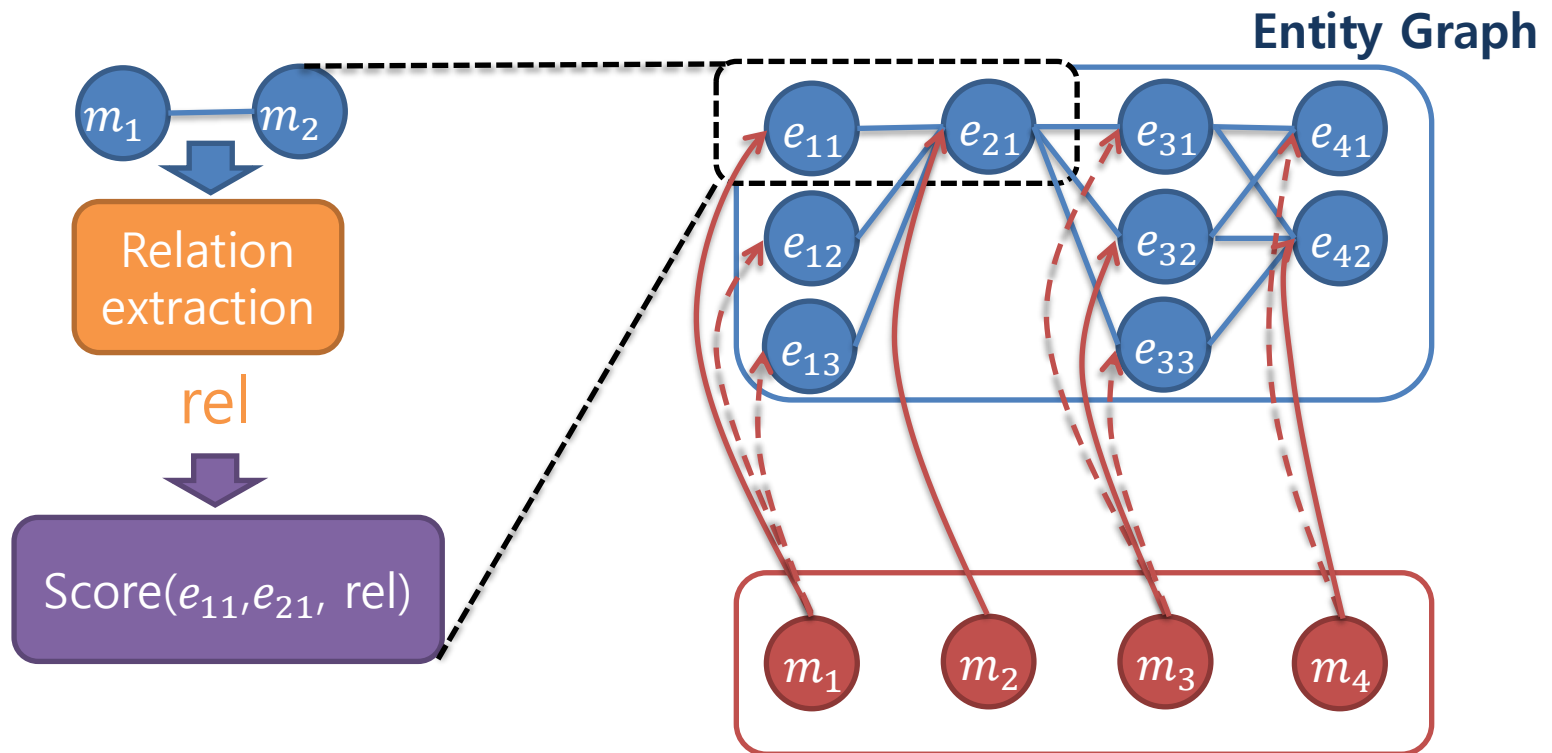


- 멘션 사전을 이용하여 가능한 모든 멘션을 탐지하고 중의성을 해결하는 모델
- 멘션 span 표상과 후보 개체 간의 유사도를 계산하여 개체 결정
- 본 연구는 위 논문의 모델을 **Global Model**로 확장하는 연구

지식 그래프를 이용한 뉴럴 엔터티 링크링

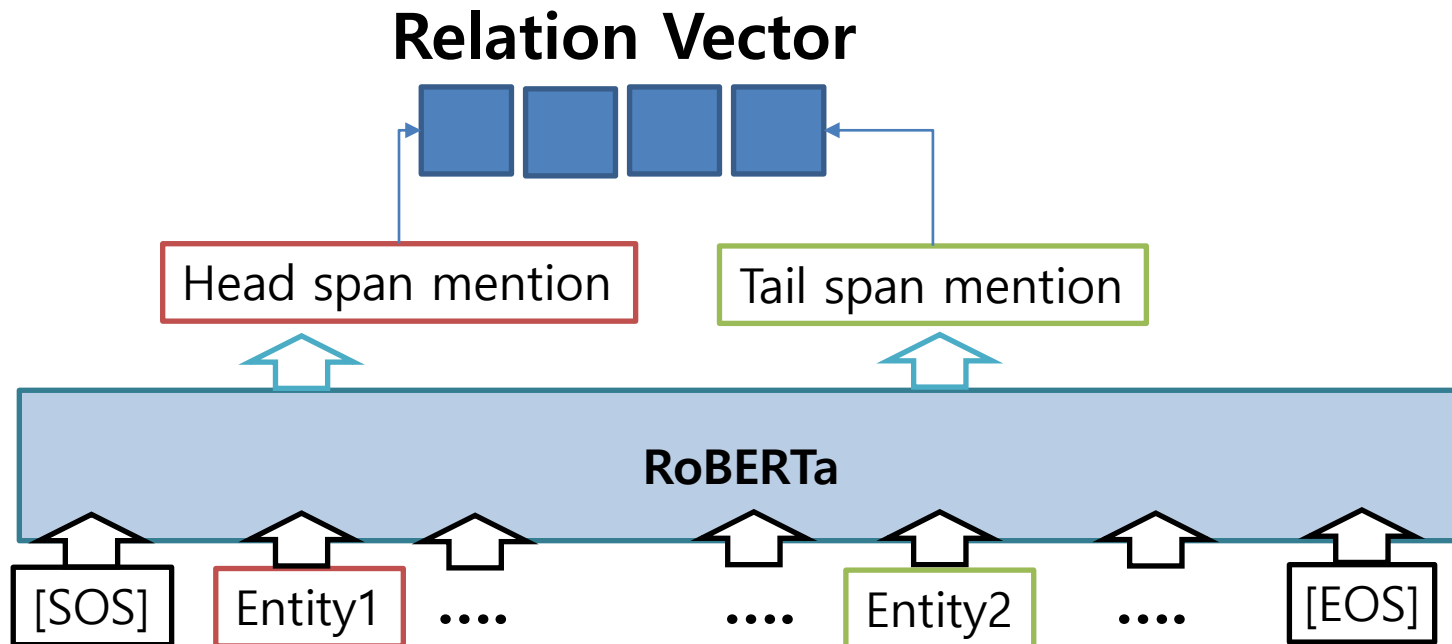
- Model 구조

- 사전 학습한 지식 그래프 임베딩을 이용해서 트리플(h,r,t) 점수를 계산
- 얻어진 점수를 기존의 GATs의 가중치로 사용하여 GATs 업데이트 진행



지식 그래프를 이용한 뉴럴 엔터티 링크링

- 지식 그래프 임베딩 사전 학습
 - 문장에 두 개체가 동시에 나타나고 인접한 두 개체는 관계 정보를 가진다고 가정
 - 나타난 두 개체에 mention span 표상을 결합하여 relation vector를 추출



지식 그래프를 이용한 뉴럴 엔터티 링크링

- 지식 그래프 임베딩 사전 학습
 - TransE와 동일한 loss함수를 사용하여 학습 진행
 - h 와 t 는 엔터티 임베딩, r 은 추출된 relation vector
 - Negative 트리플은 Positive 트리플의 h , t 을 랜덤하게 변환하여 구축

$$\mathcal{L} = \sum_{\substack{(h,r,t) \in S \\ (h',r',t') \in S'}} [Y - d(h,r,t) + d(h',r',t')]_+$$

$$[x]_+ = \max(0, x)$$

지식 그래프를 이용한 뉴럴 엔터티 링크링

- KG Score이용한 노드 업데이트
 - Graph attention network(GATs)에서 노드간 가중치를 Attention을 사용하여 결정하는데 본 연구에서는 KG score로 사용
 - $H_i^{(l)}$ 는 i번째 node 표상으로 주변 노드 $\mathcal{N}(i)$ 의 결합 정도를 $s_{ij}^{(l)}$ 를 통해 결정하여 가중합을 통해 업데이트
 - $s_{ij}^{(l)}$ 는 h+r과 t의 유사도를 통해 결정

$$H_i^{(l+1)} = g \left(W^{(l)} \sum_{j \in \mathcal{N}(i)} s_{ij}^{(l)} H_j^{(l)} + B H_i^{(l)} \right)$$

$$s_{ij} = \text{cosine_similarity}((e_i + r_{ij}), e_j)$$

지식 그래프를 이용한 뉴럴 엔터티 링크

- 개체 중의성 해결

- 멘션 span 표상과 그래프가 반영된 표상 H_{ij} 와 Biaffine 연산을 하여 최종 점수를 얻고 sorting하여 개체 결정
- 멘션 span 표상은 멘션의 start, end 표상 결합하여 구성

$$score_{ij} = Biaffine(s_i, H_{ij})$$

$$P(e_{ij}|s_i) = Softmax(score_{ij})$$

이승철_(가수)	0.65
이승철_(배우)	0.05
이승철_(기업인)	0.01

.....

실험 세팅

- 데이터셋

- 위키피디아 말뭉치를 이용하여 데이터 구성

데이터 셋	학습 셋	개발 셋	평가 셋
#Context	30000	10000	20000

... SM 아카데미 대표인 이솔림 씨의 추천으로
참가해 [거미\(가수\)](#) 거미 & 휘성의
곡인 <Do It>을 부르고

실험 세팅

- 개체 중의성 해결 모델

1. Entity Linking

베이스 모델에서 정답 멘션을 이용하여 개체 중의성만을 해결하는 모델.

2. Entity Linking+GATs

1번 모델에 GATs를 적용한 모델

3. Entity Linking+GATs+KG

1번 모델에 KG score를 적용한 모델

실험 세팅

- 엔터티 링킹 통합 모델

- 4. Joint Entity Linking

- NIL 멘션 탐지와 개체 연결의 통합 모델.

- 5. Joint Entity Linking+GATs

- 통합 모델에 GATs를 적용한 모델

- 6. Joint Entity Linking+GATs+KG

- 통합 모델에 KG score를 적용한 모델

- 7. Joint Entity Linking+GATs+KG(pipeline)

- 4번 모델을 통해 멘션 탐지를 진행하고 3번 모델을 통해 개체 중의성을 해결한 모델

실험 결과

- 개체 중의성 해결 실험

- 정답 멘션이 인식되었을 때 개체 중의성 해결 실험 결과
- 지식 그래프를 이용한 모델의 결과가 기존 모델보다 좋은 성능을 보임

모델	Linking ACC
Entity Linking	95.64%
Entity Linking+GATs	95.62%
Entity Linking+GATs+KG	96.33%

실험 결과

- 엔터티 링킹 실험 결과

- 멘션 탐지와 개체 중의성 해결을 같이 해결하는 통합 모델 실험 결과
- 통합 모델의 경우 NIL 엔터티가 존재하는데 NIL과 관련된 지식 그래프가 존재하지 않아 성능 하락

모델	멘션 탐지 F1	링킹 F1
Joint Entity Linking	89.05%	85.87%
Joint Entity Linking+GATs	89.15%	85.82%
Joint Entity Linking+GATs+KG	89.13%	85.54%
Joint Entity Linking+GATs+KG(Pipeline)	89.05%	86.42%

결론 및 향후 연구

- 결론

- 지식 그래프를 사용한 모델이 개체 중의성 해결 실험에서 기존 모델보다 향상된 성능을 보여 사전 학습한 지식 그래프가 개체간의 관계를 잘표현하여 개체명 연결에 도움이 되는 것을 확인함

- 향후 연구

- NIL 멘션 탐지와 개체 연결의 통합 모델의 경우 NIL 개체가 존재하여 지식 그래프를 반영하는데 어려움이 있어 NIL 개체 문제를 해결하는 개체 연결 통합 모델 연구

감사합니다.

Q&A