LUKE

LUKE: Deep Contextualized Entity Representations with Entity-aware Self-attention

LUKE (Language Understanding with Knowledge-based Embeddings)

- we propose new pretrained contextualized representations of words and entities based on the bidirectional transformer
- streats words and entities in a given text as independent tokens
- soutputs contextualized representations of them
- spretraining task based on the masked language model of BERT
- salso propose an entity-aware self-attention mechanism
  - sconsiders the types of tokens (words or entities) when computing attention scores

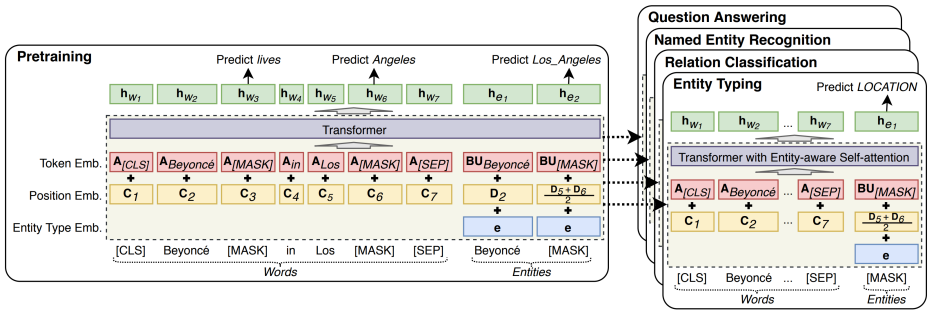
- sthe architecture of CWRs is not well suited to representing entities for the following two reasons: (1) Because CWRs do not output the span-level representations of entities, they typically need to learn how to compute such representations based on a downstream dataset that is typically small. (2) Many entity-related tasks, e.g., relation classification and QA, involve reasoning about the relationships between entities

- sAn important difference between LUKE and existing CWRs is that it treats not only words, but also entities as independent tokens, and computes intermediate and output representations for all tokens using the transformer

Contribution

- BERT 계열에서 출발.
  - 버트 + entity 토큰을 따로 학습 함
  - entity-aware self-attention mechanism
  - 토큰들의 hidden representation 뿐만 아니라 각 토큰의 entity도 학습할 수 있게 됨.

- QA, NER 등 entity의 연결관계가 필요한 task에서 SOTA 달성.



Rank	Model	EM	F1	Paper	Code	Result	Year	Tags
1	**{ANNA}** (single model)	90.622	95.719				2021	
2	**LUKE** (single model)	90.202	95.379				2020	
3	**LUKE** (single model)	90.202	95.379	LUKE: Deep Contextualized Entity Representations with Entity-aware Self-attention			2020	