Language Models are Open Knowledge Graphs[2]

Chenguang Wang[1], Xiao Liu[2], and Dawn Song[1]

[1] University of California, Berkeley
[2] Tsinghua University

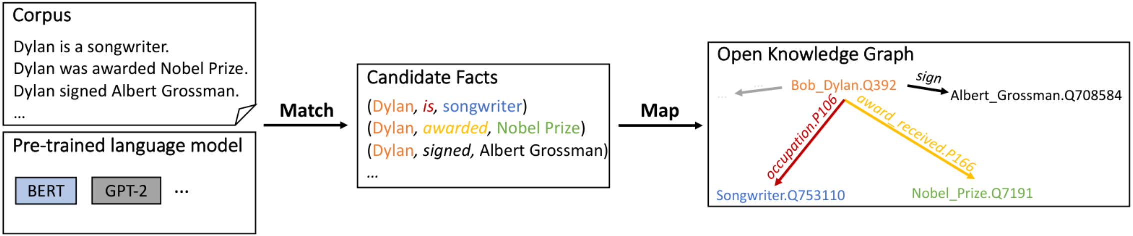
[2]Wang, Liu, and Song 2020.

Outline

Introduction

- This work presents an unsupervised approach for the task of recovery of factual knowledge stored in pre-trained language models (LMs) (GPT-2/3, BERT) to construct knowledge graphs (KGs).
- KG constructed with a single forward pass of a pre-trained LMs (no fine-tuning) over a textual corpus.
- Two-stage approach, referred as Match & Map (MaMa):
  1. **Match**: Generate candidate facts by matching of facts in the textual corpus with those in the LM.
  2. **Map**: Generates an *Open* KG by mapping of matched candidate facts to fixed as well as open KG schemas.

Match & Map (MaMa) Approach



Outline

Match Stage

- Generates candidate facts from a textual corpus.
- Leverages General and World Knowledge from large-scale corpora embedded in LMs.
- Goal: match knowledge stored in LMs with facts in the corpus.
- Facts represented as a triplet: $(head, relation, tail)$, matched via an efficient beam search in the attention weight matrices of the LM.

Beam Search

- Builds the corresponding facts between entities for every entity pair (h, t) in a sentence.
- Search maintains k-best matched candidate facts, determined by an associated matching degree computed using attention weights.
- Effectively performs one of three actions:
  1. **START**: Search from head h, adding as initial candidate.
  2. **YIELD**: Add new intermediate candidate, based on next largest attended token. 'x' marks prior tokens not considered in search.
  3. **STOP**: Search reached tail t, implying addition of candidate as a valid candidate fact.
- Algorithm is run in bidirectional manner to compensate for facts in reverse.

Beam Search

Step	Action	Intermediate candidates	Matching degrees
0	START	(Dylan,	0
1	YIELD	(Dylan, is	0.3
2	YIELD	(Dylan, is songwriter	0.7
3	STOP	(Dylan, is, songwriter)	0.7

(a) Matching example.

Query:

Key:	Dylan	is	a	songwriter
Dylan	x	x	x	x
is	0.3	x	x	x
a	0.1	0.2	x	x
songwriter	0.1	0.4	0.2	x

(b) Attention matrix for matching degree calculation.

[8]Wang, Liu, and Song 2020.

Filter

Following matching, facts of form (h, r, t) filtered as per useful constraints:

1. Matching Degree of (h, r, t) lies above a threshold.
2. Distinct frequency of r is above a threshold.
3. Relation r is a contiguous sequence in the sentence.

Map Stage

- Produces an open KG using matched candidate facts.
- Mapping to both fixed and open schema: based on existence of schema of candidate facts into an existing KG.
- Open KG comprises of mapped and unmapped facts corresponding to fixed and open schema:
  1. Mapped: $(Dylan, is, songwriter)$ mapped to $(BobDylan.Q392, occupation.P106, Songwriter.Q753110)$ as per Wikidata schema.
  2. Unmapped: $(Dylan, signed, AlbertGrossman)$ partially mapped to $(BobDylan.Q392, sign, AlbertGrossman.Q708584)$ in open schema.
- Mapping to fixed KG schema achieved by entity linking of (h, t) to (h_k, t_k) and relation mapping of (h_k, r, t_k) to (h_k, r_k, t_k).

Outline

Experiments and Results

Method	Precision %	Recall %	F1 %
OpenIE 5.1 [2]	56.98	14.54	23.16
Stanford OpenIE (Angeli et al., 2015)	61.55	17.35	27.07
MaMa-BERT$_{\text{BASE}}$ (ours)	61.57	18.79	28.79
MaMa-BERT$_{\text{LARGE}}$ (ours)	61.69	18.99	29.05
MaMa-GPT-2 (ours)	61.62	18.17	28.07
MaMa-GPT-2$_{\text{MEDIUM}}$ (ours)	62.10	18.65	28.69
MaMa-GPT-2$_{\text{LARGE}}$ (ours)	62.38	19.00	29.12
MaMa-GPT-2$_{\text{XL}}$ (ours)	62.69	19.47	**29.72**

Table 2: Compare the quality of mapped facts on TAC KBP.

Method	Precision %	Recall %	F1 %
Stanford OpenIE (Angeli et al., 2015)	23.32	13.09	16.77
MaMa-BERT$_{\text{LARGE}}$ (ours)	29.52	16.56	21.22
MaMa-GPT-2$_{\text{XL}}$ (ours)	31.32	17.42	**22.39**

Table 3: Compare the quality of mapped facts on Wikidata.

Observations

- Generation of improved KGs compared to Stanford OpenIE, with better precision but moderate recall.
- Larger/Deeper LMs produce higher quality KGs
- Significant factor of unmapped facts are due to incorrect/missing entity linking and incorrect relation mapping.

Reason	Percentage
Incorrect Entities (due to incorrect noun chunk)	33.1%
Incorrect Relation Phrases	24.8%
Correct Facts not in Reference KG	23.8%
Missing Relation Mapping	18.3%

Table: Percentage of contribution to precision errors in a sample of 100 documents

Parameter Study

- Larger beam size yields better F1 scores over held-out data.
- Larger matching degree threshold and larger distinct relation threshold decreases F1 scores.
- Attention weights of last layer work better than averaged weights.
- For reduction across multi-heads, mean reduction proved better performant.

Bibliography

Wang, Chenguang, Xiao Liu, and Dawn Song (2020). "Language Models are Open Knowledge Graphs". In: *CoRR* abs/2010.11967. arXiv: 2010.11967. URL: https://arxiv.org/abs/2010.11967.