

Lec 18

Learn From/For Knowledge Bases

Knowledge Bases — usually structured database graph with nodes & relations

Types

▷ WordNet (1995)

node: set of synonyms

edge: nouns: ← is a →

← part of →

verbs: communicate → talk → whisper

adjective: antonymy

▷ Cyc (1995) — tries to manually curate all common sense knowledge

▷ DBpedia (2007) — grab structured data from wikipedia

▷ WikiData (2008) — extract from wikipedia & many other places ↳ can search with query language

Relation Extraction

▷ Consistency in embedding

↳ word2vec shows embedding arithmetics

▷ Knowledge graph embeddings

→ Bordes + 2013 — vector stuff, triple margin loss

→ Socher + 2013 — NN to predict whether relation exists

▷ From text directly

→ Mintz + 2009 — given (entity, rel, entity), find texts with both entities in it

→ Zeng + 2014 — extract feature & classify

* Noisy data :C

→ Luo + 2017 — model the noise

↳ normalisation & transformation tricks

Using knowledge bases to inform NNs

▷ Improve embedding

→ Faruqui + — post-hoc transformation to make certain things closer

▷ Inject knowledge into LM

→ look up and prompt

→ templatic generation, with relation btwn blanks

▷ Text corpus as knowledge base

→ Relevance matching via mention vector

Schema-free Extraction

Jointly figure out schema & rel

▷ Open Info Extraction (2007)

→ No schema — the text is relation

▷ Rule-based Open IE

→ parse then check against rules

↳ some statistics to filter out noise

→ neural models

↳ dataset trick: ask simple questions about semantic role

▷ Tensor Decomposition

→ Sutskever + 2009 decompose relation & predict missing rel

▷ Universal Schema

→ Embed multiple DBs into same emb space

Relation Path Modelling

▷ Multi-step path in the graph

▷ Make the whole path differentiable

↳ certain things make certain deduction more differentiable

Probing LM's knowledge

▷ Use natural language to query KBs

Use KB as ground truth

▷ Make that multilingual — same query in different LM but LM does worse in other langs.

... so ... is this a language problem or knowledge problem?

▷ Close-book fine-tuned TS

- This helps!