

Hierarchical Losses and New Resources for Fine-grained Entity Typing and Linking

Shikhar Murty, Patrick Verga, Luke Vilnis,
Irena Radovanovic, Andrew McCallum

Xiaofan Yan

Introduction

- Model
 - Presents new methods using real and complex bilinear mappings for integrating hierarchical information, yielding substantial improvement over flat predictions in entity linking and fine-grained entity typing
- Dataset
 - MedMentions
 - A dataset identifying and linking entity mentions in PubMed abstracts to specific UMLS concepts.
 - TypeNet
 - A new dataset of hierarchical entity types for extremely fine-grained entity typing

Model

- Background
- Mention encoder
 - Converts each mention m to a d dimensional vector to classify the type or entity of the mention.
 - Embedding:
 - mention embedding: position embeddings
 - final context embedding: CNN

Training

- Mention typing: multilabel prediction
- Entity typing
 - distant supervision to noisily label mentions
 - Use multi-instance multi-label learning (MIML) to alleviate the noise from labeling
- Entity linking
 - Candidates generation: **string similarity score** *csim*
 - Compute a **similarity score** between sentence embedding and candidate entity embedding
 - Combine string cosine similarity *csim* and similarity score via a learned linear combination to generate a **final score**

Encoding Hierarchies

Experiments

- Models
 - CNN(real bilinear maps)/CNN+Complex(complex bilinear maps)/Transitive
- Mention-Level Typing in FIGER
 - Baseline: Shimaoka et al.(2017)
 - Result

Model	Acc	Macro F1	Micro F1
Ling and Weld (2012)	47.4	69.2	65.5
Shimaoka et al. (2017) †	55.6	75.1	71.7
Gupta et al. (2017) †	57.7	72.8	72.1
Shimaoka et al. (2017) ‡	59.6	78.9	75.3
CNN	57.0	75.0	72.2
+ hierarchy	58.4	76.3	73.6
CNN+Complex	57.2	75.3	72.9
+ hierarchy	59.7	78.3	75.4

Experiments

- Entity-Level Typing in TypeNet
 - Each predicted type is assigned a predicted score to be ranked to calculate MAP
 - Results:

Model	Low Data	Full Data
CNN	51.72	68.15
+ hierarchy	54.82	75.56
+ transitive	57.68	77.21
+ hierarchy + transitive	58.74	78.59
CNN+Complex	50.51	69.83
+ hierarchy	55.30	72.86
+ transitive	53.71	72.18
+ hierarchy + transitive	58.81	77.21

Experiments

- MedMentions Entity Linking with UMLS

- Results:

Model	original	normalized
mention tfidf	61.09	74.66
CNN	67.42	82.40
+ hierarchy	67.73	82.77
CNN+Complex	67.23	82.17
+ hierarchy	68.34	83.52