Towards Better Entity Linking

Mingyang Li, Yuqing Xing, Fang Kong, Guodong Zhou

Frontiers of Computer Science, DOI: 10.1007/s11704-020-0192-9

Problems & Ideas

- Problems of Entity Linking
  - For the general KBs, there normally exists a great proportion of useless (even noisy) information, which much affect the performance of entity linking, especially on the accuracy.
  - The existing entity linking model is not enough to mine text information, and can not adapt to the larger scale knowledge base.

- Ideas 1: Improve the quality of KBs
  - We propose an unsupervised KB extraction approach according to both prior knowledge and contextual relevance. In particular, a coarse-to-fine approach is employed to KB extraction.
- Ideas 2: Improve the entity model itself
  - We incorporate both the local context and the global topic information with highway network to better represent those candidates and determine the final entry.

Main Contributions

- **Coarse-to-fine KB Extraction Approach**
  - Wiki_2014_3/4_0.6 and Wiki_2018_2/3_0.8 save about 70% of the storage space and about 60% of the running time while achieving comparative performances with the original knowledge bases.

KB	Number docs	Number anchor	Size(G)	Run time(h)
Wiki_2014	4459082	18611834	11.16	0.95
Wiki_2014_3/4_0.6	**451961**	**5029896**	**3.32**	**0.42**
Wiki_2018	9618296	26916035	16.78	1.29
Wiki_2018_2/3_0.8	**682810**	**7702139**	**5.07**	**0.56**

- **Neural Entity Linking Modeling with Self-attention**
  - Our model can further capture the text information, and can better adapt to the larger scale of the knowledge base.

KB	Methods	AIDA-B
Wiki_2014	NEL-SA	92.69±0.15
Wiki_2018	NEL-SA	93.11±0.17
Wiki_merge	NEL-SA	93.22±0.19

KB	Methods	MSNBC	AQUAINT	ACE2004	CWEB	WIKI	Avg
Wiki_2014	NEL-SA	94.0±0.1	88.1±0.5	89.1±0.5	77.6±0.1	77.9±0.1	85.34
Wiki_2018	NEL-SA	94.2±0.1	88.1±0.6	89.3±0.3	77.9±0.1	77.9±0.1	85.48
Wiki_ merge	NEL-SA	94.2±0.1	87.9±0.5	89.2±0.4	77.8±0.1	78.0±0.1	85.42