

Connecting Words and Linked Data Concepts by Latent Features

Márius Šajgalík¹, Michal Barla¹, Mária Bieliková¹ and
Julian Szymański²

The Beginning of Linked Data

- easy to manually create small graphs
- a flexible structured representation of information

Linked Data Today

- big graphs
- problematic management of diverse linked data sources
 - low quality
 - duplicates
 - incomplete
 - contradictions

Unsupervised Learning of Word Features

- based on distributional hypothesis
- distributed representation of words
 - to aggregate multiple senses
 - to express multiple concepts

Our Vision - One Unified Model

- to understand raw text on the fly
 - to automatically infer mentions and descriptions of the respective linked data concepts
 - to learn new unseen concepts
 - to understand relations between concepts
 - to automatically infer new relations
 - to generate descriptions of words, concepts and relations
-
- to enable much more efficient processing of new unannotated texts

Our Vision - Almost There

We already have

- a joint model to learn latent features of words and relations [1]
- a joint model [2] of
 - words from unstructured text
 - concepts from linked data
 - relations between concepts
- everything embedded in one big latent feature space

Limitations

- static models - operates on word-level only
- limited by vocabulary

Our Vision - Realisation

- a character-wise temporal model
 - RNN used successfully in multiple works [3,4]
- limited only by the feature space, not the vocabulary
- easily handles misspellings
- easily learns the syntactical features of words

- joint dynamic model of words, concepts and relations
 - easy human interaction with big linked data

References

1. A. Bordes, J. Weston, R. Collobert, and Y. Bengio. Learning structured embeddings of knowledge bases. In AAI Conference on Artificial Intelligence, 2011.
2. J. Weston, A. Bordes, O. Yakhnenko, and N. Usunier. Connecting language and knowledge bases with embedding models for relation extraction. arXiv preprint arXiv:1307.7973, 2013.
3. J. Chung, C. Gulcehre, K. Cho, and Y. Bengio. Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555, 2014.
4. K. S. Tai, R. Socher, and C. D. Manning. Improved semantic representations from tree-structured long short-term memory networks. arXiv preprint arXiv:1503.00075, 2015.