Supervised Distributional Hypernym Discovery via Domain Adaptation

Luis Espinosa-Anke[1,3], Jose Camacho-Collados[2], Claudio Delli Bovi[2], Horacio Saggion[1]
[1]TALN Group, Department of Information and Communication Technologies
[2]Department of Computer Science, Sapienza University of Rome
[3]Savanamed
{luis.espinosa, horacio.saggion}@upf.edu, {collados,dellibovi}@di.uniroma1.it

Motivation

- ➔ The capacity for *generalization* lies at the core of human understanding.
- ➔ Lexical taxonomies are important resources on which NLP systems rely for detecting generalizations.
  - ◆ In a taxonomy learning context, the step of hypernym discovery is crucial, and a research topic in itself.
- ➔ There are two main approaches to hypernym discovery: Path/pattern based, and distributional.

Contribution

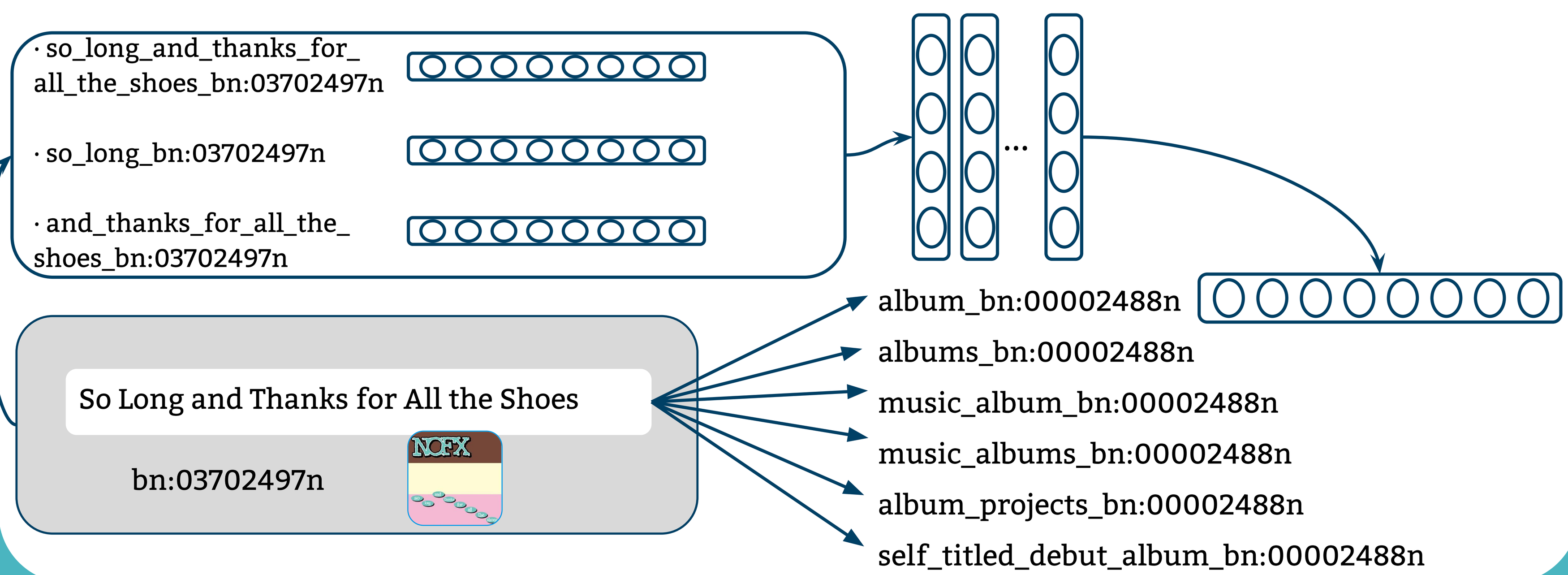
I. Break down the training data in knowledge *domains* by using the distributional approach of NASARI (Camacho-Collados et al. 2016).
II. Train a domain-wise *transformation matrix* (Mikolov et al. 2013), and use it to discover hypernyms.
III. Improve the quality of the system by incorporating disambiguated triples coming from Open Information Extraction techniques.

Training

- ➔ Obtain *is-a* sense-level **term-hypernym pairs from Wikidata**.
- ➔ **Train a transformation matrix for each domain** such that:

$$\min_{\Psi} \sum_{i=1}^{|\Phi|} \|\Psi t_i - h_i\|^2$$

- ➔ **Apply this matrix to an unseen domain-specific term**, so that the resulting vector constitutes the "ideal" hypernym for that term. Since it may not coincide with any predefined vector, retrieve its nearest neighbours by cosine similarity.



Resources

- ➔ **BabelNet** (Navigli and Ponzetto, 2012) - The largest multilingual repository of concepts and entities.
- ➔ **SensEmbed** (Iacobacci et al. 2015) - A sense-level real-valued vector space representation, where each vector corresponds to a BabelNet ***synset*** and its ***lexicalization***.
  - ◆ E.g. v(***bass_bn:00008917n***) = [0.2346, -0.756222, 0.123236 …]
- ➔ **KB-Unify** (Delli Bovi et al. 2015) - An integration of Open Information Extraction systems, disambiguated using BabelNet as reference sense inventory. It contains triples from Patty, WiseNet, NELL and ReVerb.

Conclusion

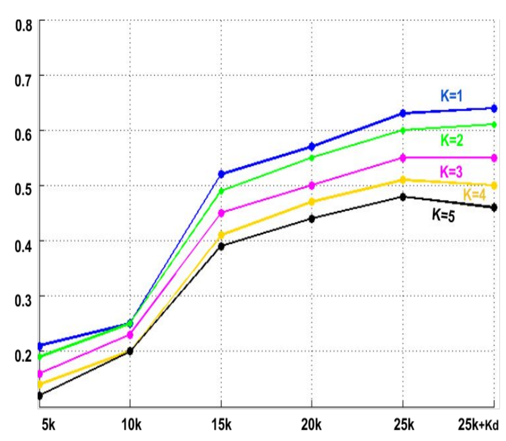
We perform experiments on hypernym *discovery*. Traditionally, systems are evaluated either on detecting a hypernymic relation in a pair of concepts, or in finding the best hypernym from a predefined and closed terminology. Providing a hypernym *from scratch* and link it to a knowledge resource is more challenging. **Key findings:**

- ➔ **Domain clustering is essential**. This is consistent with the intuition of Fu et al. (2014).
- ➔ **In some domains, feeding OIE** triples to the training data **improves**, *but not always*.

Hypernym Discovery Evaluation

Train	education			biology			transport		
	MRR	MAP	R-P	MRR	MAP	R-P	MRR	MAP	R-P
5k	0.00	0.00	0.00	0.63	0.63	0.59	0.01	0.01	0.01
15k	0.22	0.22	0.21	**0.84**	0.72	0.79	0.25	0.23	0.21
25k	0.33	0.32	0.30	**0.84**	**0.83**	**0.81**	0.46	0.43	0.39
25k+KBU$_{25k}$	**0.38**	**0.36**	**0.33**	0.70	0.63	0.56	**0.48**	**0.45**	**0.41**
100k Random	0.00	0.00	0.00	**0.84**	0.81	0.77	0.01	0.02	0.02
Baseline	0.10	0.10	0.09	0.58	0.57	0.57	0.29	0.25	0.21

P@K- Transport



Results for other seven domains available in the paper.

Extra-Coverage

Manual evaluation **outside of Wikidata**:

- Three pattern-based comparison systems: **Yago, WiBi** and **DefIE**.
- Precision lower than these approaches but **competitive recall**.
- Interesting follow-up in **combining our model with pattern-based systems**, in the line of Shwartz et al. (2016).

Data & Code

- BabelNet synsets clustered by domain.
- Wikidata and KBU *isa* branches.
- Python API
  - Word, synset and sense level.
  - Batch predict and interactive console.

taln.upf.edu/taxoembed

References

[1] Camacho-Collados, J., Pilehvar, M.T., & Navigli, R. (2016). Nasari: Integrating explicit knowledge and corpus statistics for a multilingual representation of concepts and entities. AIJ, 240, 36-64.

[1] Navigli, R., & Ponzetto, S. P. (2012). BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. AIJ, 193, 217-250.

[2] Iacobacci, I., Pilehvar, M. T., & Navigli, R. (2015). SensEmbed: learning sense embeddings for word and relational similarity. In Proceedings of ACL (pp. 95-105).

[3] Delli Bovi, C., Espinosa-Anke, L., & Navigli, R. (2015). Knowledge Base Unification via Sense Embeddings and Disambiguation. In Proceedings of EMNLP (pp. 726-736).

[4] Mikolov, T., Le, Q. V., & Sutskever, I. (2013). Exploiting similarities among languages for machine translation. arXiv preprint arXiv:1309.4168.

[5] Fu, R., Guo, J., Qin, B., Che, W., Wang, H., & Liu, T. (2014). Learning Semantic Hierarchies via Word Embeddings. In Proceedings of ACL (pp. 1199-1209).

[6] Shwartz, V. Goldberg, & Dagan, I.. (2016). Improving Hypernymy Detection with an Integrated Path-based and Distributional Method. In Proceedings of ACL (pp. 2389-2398).

EXCELENCIA MARÍA DE MAEZTU
SAPIENZA UNIVERSITÀ DI ROMA
SAVANA