Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

Alexander Panchenko, Dmitry Ustalov,
Stefano Faralli, Simone Paolo Ponzetto, and
Chris Biemann

IMPROVING HYPERNYMY EXTRACTION WITH DISTRIBUTIONAL SEMANTIC CLASSES

Introduction

Hypernyms

Examples of hypernymy relations

- **apple** –isa→ **fruit**
- **mangosteen** –isa→ **fruit**

Hypernyms

Examples of hypernymy relations

- **apple#1** –isa→ **fruit#2**
- **mangosteen#0** –isa→ **fruit#2**

Hypernyms

Examples of hypernymy relations

- **apple#1** –isa→ **fruit#2**
- **mangosteen#0** –isa→ **fruit#2**

"This café serves fresh mangosteen juice"

Hypernyms

Examples of hypernymy relations

- **apple#1** –isa→ **fruit#2**
- **mangosteen#0** –isa→ **fruit#2**

“This café serves fresh mangosteen juice”

Examples of applications of hypernyms

- question answering [Zhou et al., 2013]
- query expansion [Gong et al., 2005]
- semantic role labelling [Shi & Mihalcea, 2005]

Automatic extraction of hypernyms

A short history of extraction methods

1. [Hearst, 1992]: lexical-syntactic patterns **defined manually**;

Automatic extraction of hypernyms

A short history of extraction methods

1. [Hearst, 1992]: lexical-syntactic patterns **defined manually**;
2. [Snow et al., 2004]: lexical-syntactic patterns learned in **a supervised way**;

Automatic extraction of hypernyms

A short history of extraction methods

1. [Hearst, 1992]: lexical-syntactic patterns **defined manually**;
2. [Snow et al., 2004]: lexical-syntactic patterns learned in **a supervised way**;
3. [Weeds et al., 2014]: supervised approach with **word embedding features**;
4. [Shwartz et al., 2016]: supervised approach with **word and path embedding features**;
5. [Glavaš & Ponzetto, 2017, Ustalov et al., 2017]: taking into account **assymetry** of hypernyms.

Automatic extraction of hypernyms

A short history of extraction methods

1. [Hearst, 1992]: lexical-syntactic patterns **defined manually**;
2. [Snow et al., 2004]: lexical-syntactic patterns learned in **a supervised way**;
3. [Weeds et al., 2014]: supervised approach with **word embedding features**;
4. [Shwartz et al., 2016]: supervised approach with **word and path embedding features**;
5. [Glavaš & Ponzetto, 2017, Ustalov et al., 2017]: taking into account **assymetry** of hypernyms.

Not taking into account word senses and global structure!

Induction of semantic classes

"Global distributional structure" of a language $\approx$ global sense clustering.

Induction of semantic classes

"Global distributional structure" of a language ≈ global sense clustering.

A short history of extraction methods

1. [Lin & Pantel, 2001]: sets of similar words are clustered into concepts.
2. [Pantel & Lin, 2002]: words can belong to several clusters (representing senses)
3. [Pantel & Ravichandran, 2004]: aggregate hypernyms per cluster from from Hearst patterns

No explicit evaluation of usefulness of hypernymy labels for hypernymy extraction.

Main contributions

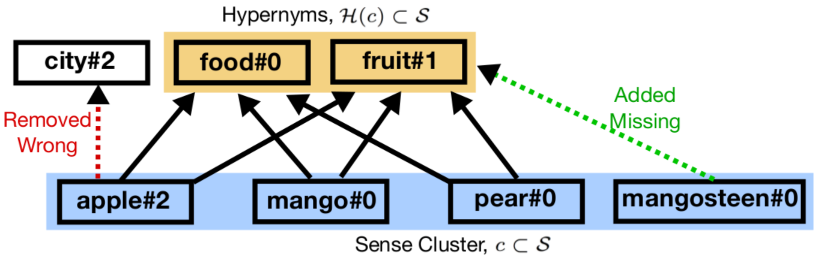
- We show how distributionally-induced **semantic classes** can be helpful for **extracting hypernyms**:

Main contributions

- We show how distributionally-induced **semantic classes** can be helpful for **extracting hypernyms**:
  1. A method for **inducing sense-aware semantic classes** using distributional semantics;
  2. A method for using the induced semantic classes for **filtering noisy hypernymy relations**.

Method

Labeled semantic classes



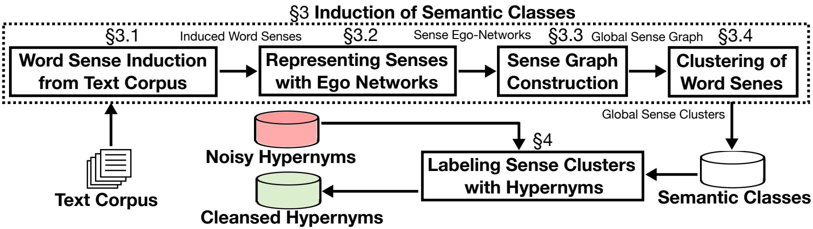
- **Post-processing of hypernymy relations** using distributionally induced semantic classes;
- A semantic class is a clusters of induced word senses **labeled with hypernyms**.

Outline of our approach

1. Sense-aware distributional semantic classes are **induced from a text corpus**;
2. Semantic classes are used to **filter a noisy hypernyms** database.

Outline of our approach

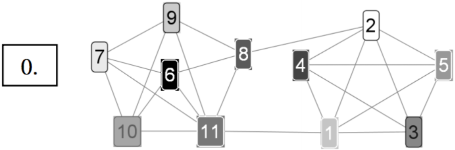
1. Sense-aware distributional semantic classes are **induced from a text corpus**;
2. Semantic classes are used to **filter a noisy hypernyms** database.



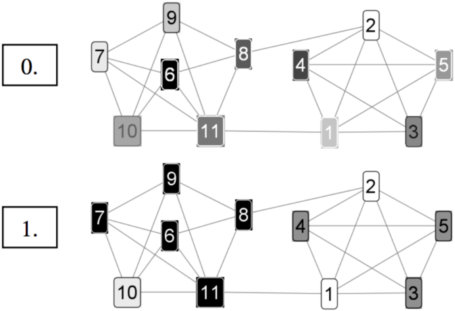
Chinese Whispers#1

* source of the image: http://ic.pics.livejournal.com/blagin_anton/33716210/2701748/2701748_800.jpg

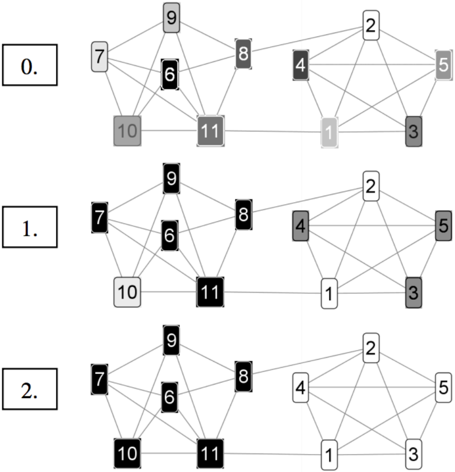
Chinese Whispers#2: graph clustering



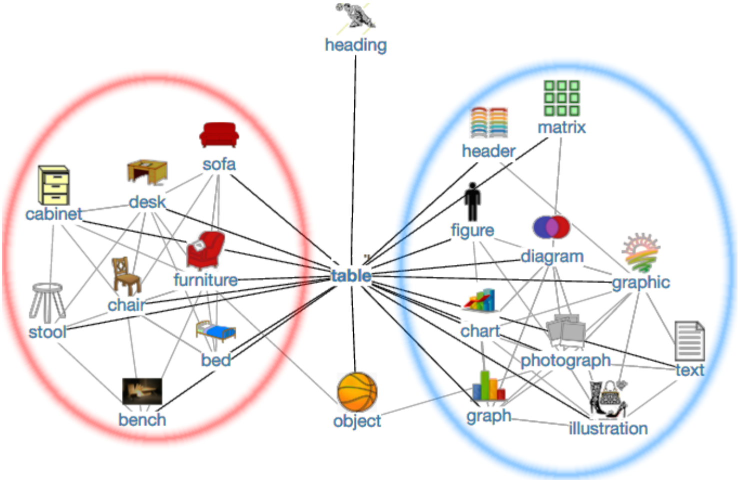
Related work: Chinese Whispers#2



Related work: Chinese Whispers#2



Graph-based word sense induction



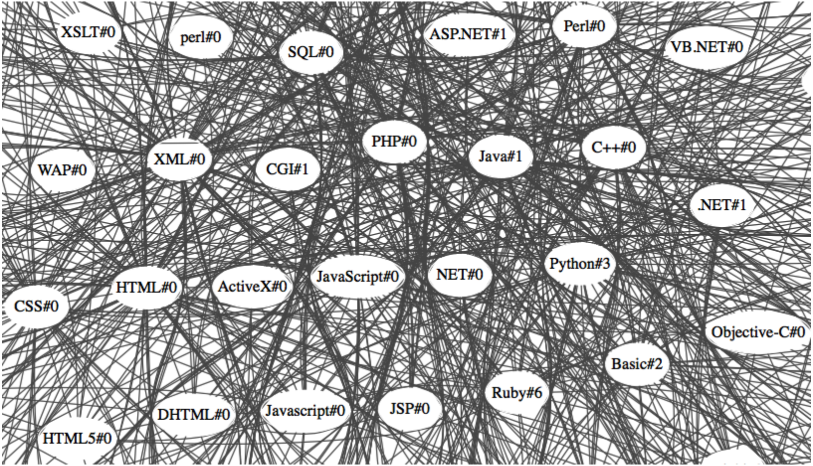
Sample of induced sense inventory

Word Sense	Local Sense Cluster: Related Senses	Hypernyms
mango#0	peach#1, grape#0, plum#0, apple#0, apricot#0, watermelon#1, banana#1, coconut#0, pear#0, fig#0, melon#0, **mangosteen#0**, ...	fruit#0, food#0, ...
apple#0	mango#0, pineapple#0, banana#1, melon#0, grape#0, peach#1, watermelon#1, apricot#0, cranberry#0, pumpkin#0, **mangosteen#0**, ...	fruit#0, crop#0, ...
Java#1	C#4, Python#3, Apache#3, Ruby#6, Flash#1, C++#0, SQL#0, ASP#2, Visual Basic#1, CSS#0, Delphi#2, MySQL#0, Excel#0, Pascal#0, ...	programming language#3, language#0, ...
Python#3	PHP#0, Pascal#0, Java#1, SQL#0, Visual Basic#1, C++#0, JavaScript#0, Apache#3, Haskell#5, .NET#1, C#4, SQL Server#0, ...	language#0, technology#0, ...

Sample of induced semantic classes

ID	Global Sense Cluster: Semantic Class	Hypernyms
1	peach#1, banana#1, pineapple#0, berry#0, blackberry#0, grapefruit#0, strawberry#0, blueberry#0, mango#0, grape#0, melon#0, orange#0, pear#0, plum#0, raspberry#0, watermelon#0, apple#0, apricot#0, watermelon#0, pumpkin#0, berry#0, **mangosteen#0**, ...	vegetable#0, fruit#0, crop#0, ingredient#0, food#0, ·
2	C#4, Basic#2, Haskell#5, Flash#1, Java#1, Pascal#0, Ruby#6, PHP#0, Ada#1, Oracle#3, Python#3, Apache#3, Visual Basic#1, ASP#2, Delphi#2, SQL Server#0, CSS#0, AJAX#0, JavaScript#0, SQL Server#0, Apache#3, Delphi#2, Haskell#5, .NET#1, CSS#0, ...	programming language#3, technology#0, language#0, format#2, app#0

Network of induced word senses



Optimization of meta-parameters

Comparison to WordNet and BabelNet

Meta-parameters

1. **Min. num. of sense co-occurrences in an ego-network**: $t > 0$
2. **Sense edge weight type**: count or log(count)
3. **Hypernym weight type**: tf-idf or tf

Comparison to WordNet and BabelNet

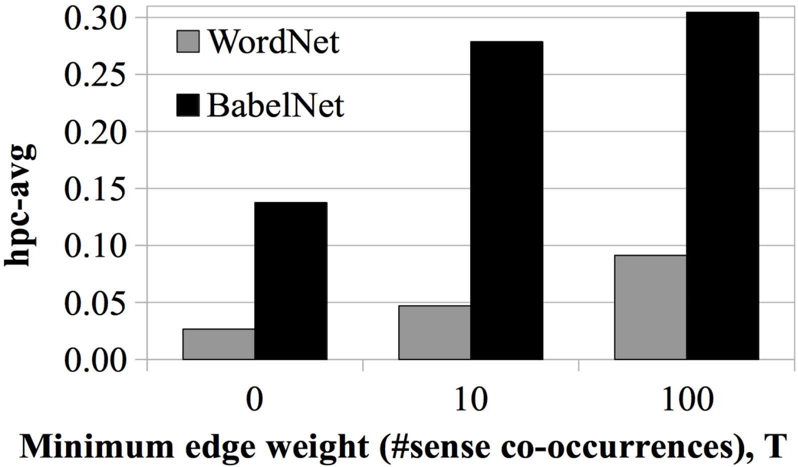
Meta-parameters

1. **Min. num. of sense co-occurrences in an ego-network**: $t > 0$
2. **Sense edge weight type**: count or log(count)
3. **Hypernym weight type**: tf-idf or tf

$$hpc\text{-score}(c) = \frac{h\text{-score}(c) + 1}{p\text{-score}(c) + 1} \cdot \text{coverage}(c).$$

$$p\text{-score}(c) = \frac{1}{|c|} \sum_{i=1}^{|c|} \sum_{j=1}^{i} \text{dist}(w_i, w_j). \; h\text{-score}(c) = \frac{|\mathcal{H}(c) \cap \text{gold}(c)|}{|\mathcal{H}(c)|}.$$

Impact of the min. edge weight *t*

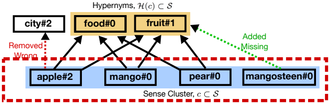


Best coarse- and fine-grained models

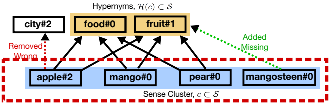
Min. num of sense co-occurr., *t*	**Edge weight**, *E*	**Hypernym weight**, *H*	**Number of clusters**	**Number of senses**	*hpc*-avg, **WordNet**	*hpc*-avg, **BabelNet**
0	count	tf-idf	**1 870**	**208 871**	0.041	0.279
100	log	tf-idf	**734**	**18 028**	0.092	0.304

Results

Plausibility of Semantic Classes



Plausibility of Semantic Classes



- Layout of the sense cluster evaluation crowdsourcing task;
- the entry "**winchester**" is the **intruder**.

Topics:
- vegetable
- fruit
- crop

For these topics we have the list of the following words:
- peach
- pineapple
- winchester
- watermelon
- cherry
- blackberry

Select the words that are non-relevant for the topics above:

☐ peach
☐ pineapple
☐ winchester
☐ watermelon
☐ cherry
☐ blackberry

Plausibility of Semantic Classes

1. **Accuracy** is the fraction of tasks where annotators correctly identified the intruder;
2. **Badness**: is the fraction of tasks for which non-intruder words were selected.

Plausibility of Semantic Classes

1. **Accuracy** is the fraction of tasks where annotators correctly identified the intruder;
2. **Badness**: is the fraction of tasks for which non-intruder words were selected.

	Accuracy	**Badness**	**Randolph κ**
Sense clusters, c	0.859	0.248	0.739
Hyper. labels, $\mathcal{H}(c)$	0.919	0.208	0.705

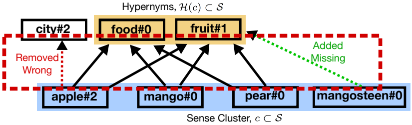
Plausibility of Semantic Classes

1. **Accuracy** is the fraction of tasks where annotators correctly identified the intruder;
2. **Badness**: is the fraction of tasks for which non-intruder words were selected.

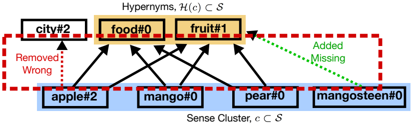
	Accuracy	**Badness**	**Randolph κ**
Sense clusters, c	0.859	0.248	0.739
Hyper. labels, $\mathcal{H}(c)$	0.919	0.208	0.705

- **Clusters**: 68 annotators, 2,035 judgments;
- **Hypernyms**: 98 annotators, 2,245 judgments.

Improving Hypernymy Relations



Improving Hypernymy Relations



Layout of the hypernymy annotation task:

Is it correct that **peach** is a kind of **fruit**?

Your opinion:

◯ Yes

◯ No

Improving Hypernymy Relations

Evaluating results of post-processing of a noisy hypernymy database using human judgements:

- A random sample of 4,870 relations using lexical split;
- each labeled 6.9 times on average;
- a total of 33,719 judgments.

Improving Hypernymy Relations

Evaluating results of post-processing of a noisy hypernymy database using human judgements:

- A random sample of 4,870 relations using lexical split;
- each labeled 6.9 times on average;
- a total of 33,719 judgments.

	Precision	**Recall**	**F-score**
Original hypernymy relations extracted from Common Crawl corpus [Seitner et al., 2016]	0.475	0.546	0.508
Enhanced hypernyms with the *coarse-grained* semantic classes	**0.541**	**0.679**	**0.602**

Improving Taxonomy Induction

- **SemEval 2016 Task 13** ”Taxonomy Extraction from Text”;
- **Fowlkes&Mallows Measure (F&M)** – a cumulative measure of the similarity of taxonomies;
- **English** part of the dataset.

Improving Taxonomy Induction

- **SemEval 2016 Task 13** "Taxonomy Extraction from Text";
- **Fowlkes&Mallows Measure (F&M)** – a cumulative measure of the similarity of taxonomies;
- **English** part of the dataset.

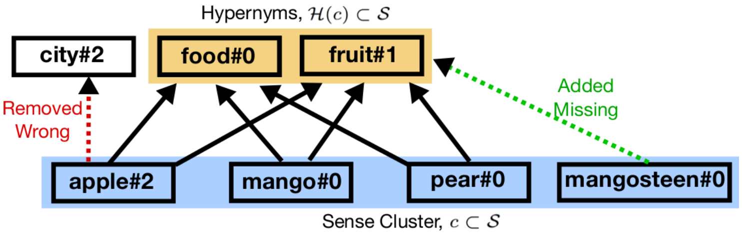
Domain	**#Seeds words**	**#Expanded words**	**#Clusters,** fine-gr.	**#Clusters,** coarse-gr.
Food	2 834	3 047	29	21
Science	806	1 137	73	35
Environ.	261	909	111	39

Improving Taxonomy Induction

System / Dataset	Food, WordNet	Science, WordNet	Food, Combined	Science, Combined	Science, Eurovoc	Environ., Eurovoc
WordNet	1.0000	1.0000	0.5870	0.5760	0.6243	n.a.
Baseline	0.0022	0.0016	0.0019	0.0163	0.0056	0.0000
JUNLP	0.1925	0.0494	0.2608	0.1774	0.1373	0.0814
NUIG-UNLP	n.a.	0.0027	n.a.	0.0090	0.1517	0.0007
QASSIT	n.a.	0.2255	n.a.	0.5757	0.3893	0.4349
TAXI	0.3260	0.2255	0.2021	0.3634	0.3893	0.2384
USAAR	0.0021	0.0008	0.0000	0.0020	0.0023	0.0007
Sem. Class, fine-gr.	0.4540	0.4181	0.5147	0.6359	**0.5831**	0.5600
Sem. Class, coarse-gr.	**0.4774**	**0.5927**	**0.5799**	**0.6539**	0.5515	**0.6326**

Summary

1. An unsupervised method for the induction of **sense-aware distributional semantic classes**;
2. Showed how these can be used for **post-processing of noisy hypernymy databases** extracted from text.



Thank you! Questions?

Glavaš, G. & Ponzetto, S. P. (2017).
Dual tensor model for detecting asymmetric lexico-semantic relations.
In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing* (pp. 1758–1768). Copenhagen, Denmark: Association for Computational Linguistics.

Gong, Z., Cheang, C. W., & Leong Hou, U. (2005).
Web Query Expansion by WordNet.
In *Proceedings of the 16th International Conference on Database and Expert Systems Applications - DEXA '05* (pp. 166–175). Copenhagen, Denmark: Springer Berlin Heidelberg.

Hearst, M. A. (1992).
Automatic Acquisition of Hyponyms from Large Text Corpora.
In *Proceedings of the 14th Conference on Computational Linguistics - Volume 2*, COLING '92 (pp. 539–545). Nantes, France: Association for Computational Linguistics.

Lin, D. & Pantel, P. (2001).

Induction of Semantic Classes from Natural Language Text.
In *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '01 (pp. 317–322). San Francisco, CA, USA: ACM.

Pantel, P. & Lin, D. (2002).
Discovering Word Senses from Text.
In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '02 (pp. 613–619). Edmonton, AB, Canada: ACM.

Pantel, P. & Ravichandran, D. (2004).
Automatically Labeling Semantic Classes.
In *Proceedings of the Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL'2004)* (pp. 321–328). Boston, MA, USA: Association for Computational Linguistics.

Seitner, J., Bizer, C., Eckert, K., Faralli, S., Meusel, R., Paulheim, H., & Ponzetto, S. P. (2016).

A Large DataBase of Hypernymy Relations Extracted from the Web.
In *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, LREC 2016 (pp. 360–367). Portorož, Slovenia: European Language Resources Association (ELRA).

Shi, L. & Mihalcea, R. (2005).
Putting Pieces Together: Combining FrameNet, VerbNet and WordNet for Robust Semantic Parsing.
In *Proceedings of the 6th International Conference on Computational Linguistics and Intelligent Text Processing*, CICLing 2005 (pp. 100–111). Mexico City, Mexico: Springer Berlin Heidelberg.

Shwartz, V., Goldberg, Y., & Dagan, I. (2016).
Improving Hypernymy Detection with an Integrated Path-based and Distributional Method.
In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp.

2389–2398). Berlin, Germany: Association for Computational Linguistics.

Snow, R., Jurafsky, D., & Ng, A. Y. (2004).
Learning Syntactic Patterns for Automatic Hypernym Discovery.
In *Proceedings of the 17th International Conference on Neural Information Processing Systems*, NIPS'04 (pp. 1297–1304). Vancouver, BC, Canada: MIT Press.

Ustalov, D., Arefyev, N., Biemann, C., & Panchenko, A. (2017).
Negative sampling improves hypernymy extraction based on projection learning.
In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers* (pp. 543–550). Valencia, Spain: Association for Computational Linguistics.

Weeds, J., Clarke, D., Reffin, J., Weir, D. J., & Keller, B. (2014).
Learning to distinguish hypernyms and co-hyponyms.

In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers* (pp. 2249–2259). Dublin, Ireland: Dublin City University and Association for Computational Linguistics.

Zhou, G., Liu, Y., Liu, F., Zeng, D., & Zhao, J. (2013).
Improving question retrieval in community question answering using world knowledge.
In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, IJCAI '13 (pp. 2239–2245). Beijing, China: AAAI Press.