Automatic Selection of Context Configurations for Improved Class-Specific Word Representations

Ivan Vulić, **Roy Schwartz**, Ari Rappoport,
Roi Reichart and Anna Korhonen

Background
Distributional Semantics: What is a Context?

The nice people rode their horses bravely and rapidly
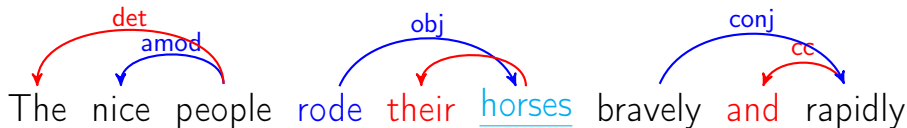
Bag-of-words

The nice people rode their horses bravely and rapidly

- ▶ Bag-of-words: simplest approach
    - ▶ Noisy

Background

Distributional Semantics: What is a Context?

- Bag-of-words: simplest approach
  - Noisy
- Dependency links: more accurate contexts
  - Are all dependency links useful for representing words?
  - Different dependency links represent different word classes

[Lin, 1998, Levy and Goldberg, 2014]

Background
Distributional Semantics: What is a Context?

conj

The nice people rode their horses bravely and rapidly
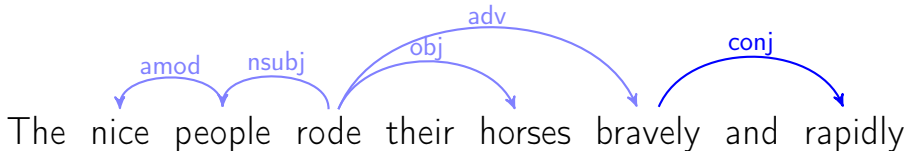
- ▶ Bag-of-words: simplest approach
    - ▶ Noisy
- ▶ Dependency links: more accurate contexts
    - ▶ Are all dependency links useful for representing words?
    - ▶ Different dependency links represent different word classes
- ▶ Coordinations / symmetric patterns: more **accurate** and more **efficient**

[Schwartz et al., 2015, Schwartz et al., 2016]

Background
Distributional Semantics: What is a Context?

The nice people rode their horses bravely and rapidly

- Bag-of-words: simplest approach
  - Noisy
- Dependency links: more accurate contexts
  - Are all dependency links useful for representing words?
  - Different dependency links represent different word classes
- Coordinations / symmetric patterns: more **accurate** and more **efficient**
  - But... valuable information gets lost

[Schwartz et al., 2015, Schwartz et al., 2016]

Main Contributions

- Detect which **fine-grained context types** are useful for different word classes
- Traverse the large space of **context configurations** efficiently to find the best context configuration
- **Transfer** the configurations learned for one task and one language **to other tasks and languages without re-training**
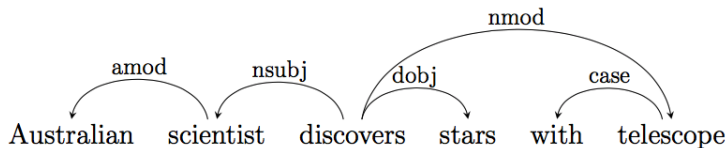
Context Types
(Universal) Labeled Dependency Edges
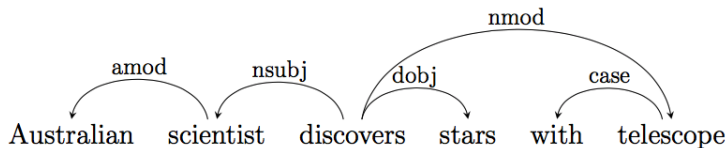
- (discovers, scientist_nsubj)
- (discovers, stars_dobj)
- (discovers, telescope_nmod)
- (stars, discovers_dobj-1)
- ...

Context Types
(Universal) Labeled Dependency Edges

- (discovers, scientist_nsubj)
- (discovers, stars_dobj)
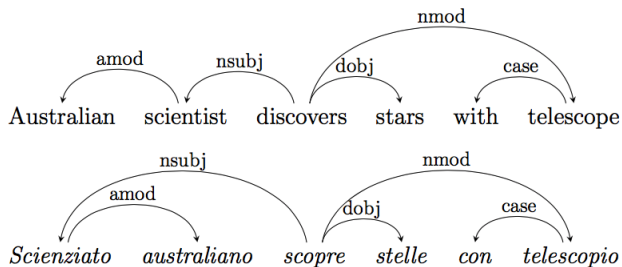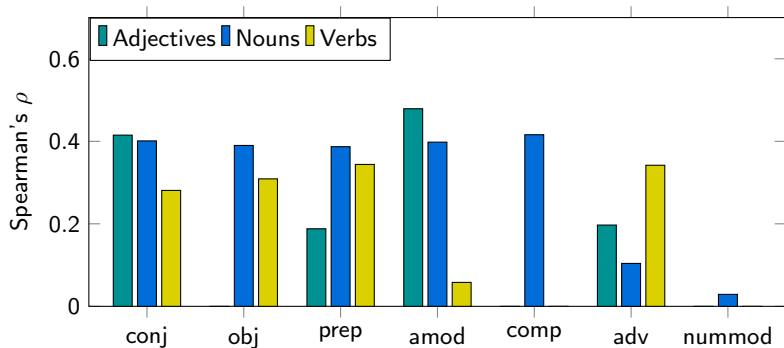- (discovers, telescope_nmod)
- (stars, discovers_dobj-1)
- ...

Cross Lingual Context Transfer?

Results: Individual Labels

Too many Context Configurations

Adjectives	Verbs	Nouns
amod, conjlr, conjll	prep, acl, obj, comp, adv, conjlr, conjll	amod, prep, comp, subj, obj, appos, acl, nmod, conjlr, conjll

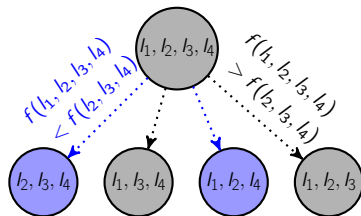- ▶ Traversing a potentially huge context configuration may be intractable

Searching for Context Configurations
An Adapted Beam-Search Algorithm

l_1, l_2, l_3, l_4
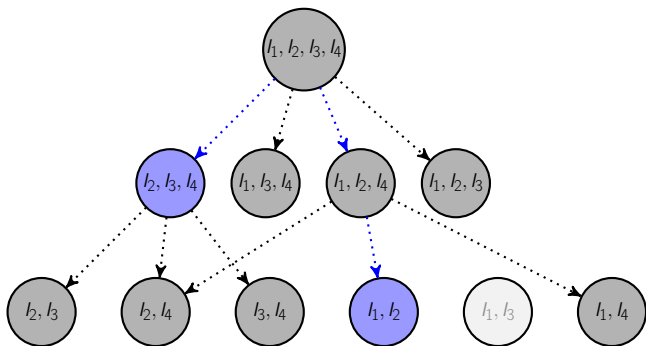
Searching for Context Configurations
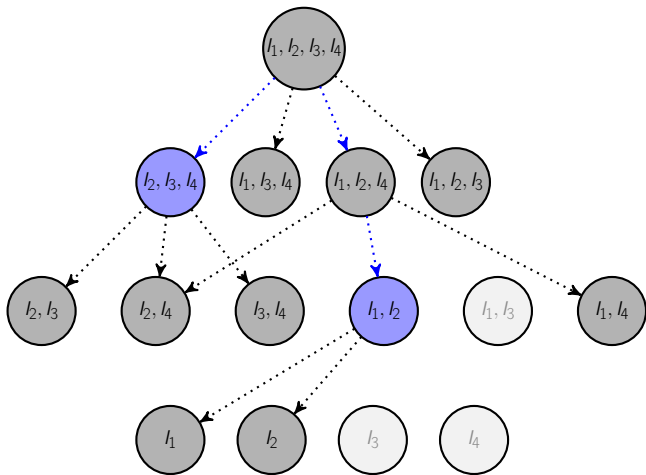## An Adapted Beam-Search Algorithm

$f(x)$: dev set evaluation

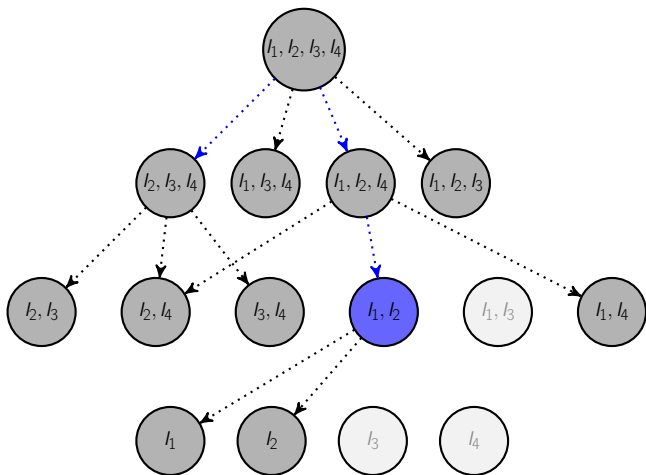# Searching for Context Configurations
An Adapted Beam-Search Algorithm

Searching for Context Configurations
An Adapted Beam-Search Algorithm

Searching for Context Configurations
An Adapted Beam-Search Algorithm

Experimental Setup

- **Model:** Skip-gram with negative sampling [Mikolov et al., 2013]

- **Training data:** Polyglot Wikipedia

- **Evaluation:** SimLex-999 word similarity dataset [Hill et al., 2015]
  - 666 noun pairs, 222 verb pairs, 111 adjective pairs
  - 2-fold cross validation
  - Evaluation measure: Spearman's ρ

- **Baselines:** A variety of standard context types
  - Bag-of-words (w/ and w/o positions); all dependency links, coordination dependency links, symmetric patterns

Results: Context Configurations
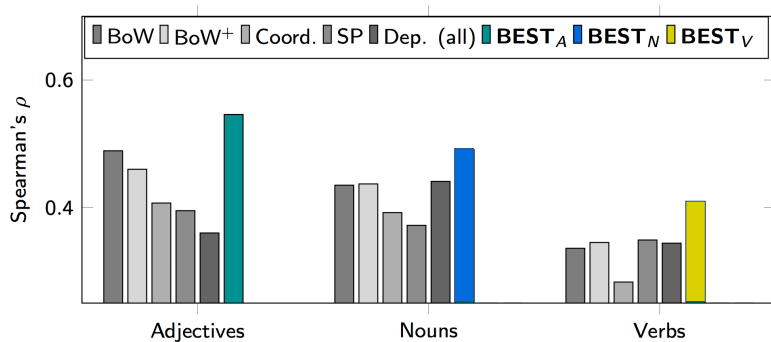
Selected Contexts are Efficient
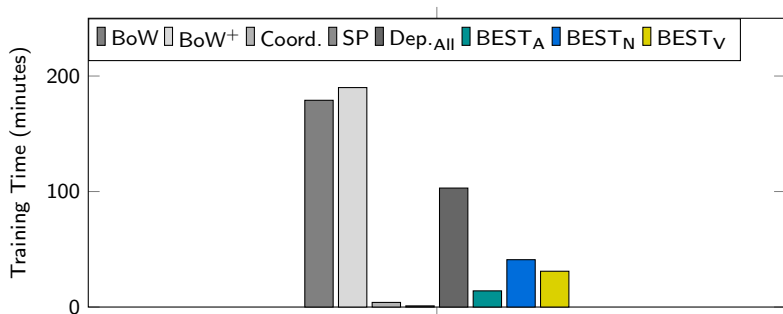
Transfer Results

- TOEFL
  - 5% improvement over strongest baseline on verbs and nouns
- Other languages
  - 0.02—0.08 ρ improvement on Italian and German accros all three word classes
    - DE and IT SimLex999 [Leviant and Reichart, 2015]

Take-Home Messages

- Different word classes require different (**finer-grained**) context configurations

- An automatic framework for **computationally tractable** selection of optimal context configurations

- Design based on Universal Dependencies: context configurations **transferable** to other tasks and languages without retraining

- **Future work** → finer-grained contexts, other word classes, more sophisticated search algorithms, other representation models, context weighting, ...

Take-Home Messages

- Different word classes require different (**finer-grained**) context configurations

- An automatic framework for **computationally tractable** selection of optimal context configurations

- Design based on Universal Dependencies: context configurations **transferable** to other tasks and languages without retraining

- **Future work** → finer-grained contexts, other word classes, more sophisticated search algorithms, other representation models, context weighting, ...

Thank you!

References I

Hill, F., Reichart, R., and Korhonen, A. (2015).
Simlex-999: Evaluating semantic models with (genuine) similarity estimation.
Computational Linguistics.

Leviant, I. and Reichart, R. (2015).
Judgment language matters: Multilingual vector space models for judgment language aware lexical semantics.
arxiv:1508.00106.

Levy, O. and Goldberg, Y. (2014).
Dependency-based word embeddings.
In *Proc. of ACL.*

Lin, D. (1998).
Automatic retrieval and clustering of similar words.
In *Proc. of ACL.*

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013).
Efficient estimation of word representations in vector space.
arXiv:1301.3781.

Schwartz, R., Reichart, R., and Rappoport, A. (2015).
Symmetric pattern based word embeddings for improved word similarity prediction.
In *Proc. of CoNLL.*

References II

Schwartz, R., Reichart, R., and Rappoport, A. (2016).
Symmetric patterns and coordinations: Fast and enhanced representations of
verbs and adjectives.
In *Proc. of NAACL.*