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

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

CoNLL 2017; Vancouver; August 3, 2017
The nice people rode their horses bravely and rapidly
The nice people rode their horses bravely and rapidly

- Bag-of-words: simplest approach
  - Noisy
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

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

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

Australian scientist discovers stars with telescope

- amod
- nsubj
- dobj
- nmod
- case
Context Types

(Universal) Labeled Dependency Edges

- (discovers, scientist_nsubj)
- (discovers, stars_dobj)
- (discovers, telescope_nmod)
- (stars, discovers_dobj-1)
- ...
Cross Lingual Context Transfer?

Australian scientist discovers stars with telescope

Scienziato australiano scopre stelle con telescopio
Results: Individual Labels

![Bar chart showing Spearman's \( \rho \) for different parts of speech categories: conj, obj, prep, amod, comp, adv, nummod.](chart.png)
Too many Context Configurations

<table>
<thead>
<tr>
<th>Adjectives</th>
<th>Verbs</th>
<th>Nouns</th>
</tr>
</thead>
<tbody>
<tr>
<td>amod,</td>
<td>prep, acl,</td>
<td>amod, prep, comp, subj, obj,</td>
</tr>
<tr>
<td>conjlr,</td>
<td>obj, comp, adv,</td>
<td>appos, acl, nmod, conjlr,</td>
</tr>
<tr>
<td>conjll</td>
<td>conjlr, conjll</td>
<td>conjll</td>
</tr>
</tbody>
</table>

- 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

\[ f(x) : \text{dev set evaluation} \]
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 $\rho$
- **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

![Diagram showing Spearman's $\rho$ for different context configurations. The x-axis represents Adjectives, Nouns, and Verbs, and the y-axis represents Spearman's $\rho$. The configurations include BoW, BoW+, Coord., SP, Dep., (all), BEST_A, BEST_N, and BEST_V.]
Selected Contexts are Efficient
Transfer Results

- **TOEFL**
  - 5% improvement over strongest baseline on verbs and nouns

- **Other languages**
  - 0.02–0.08 $\rho$ improvement on Italian and German across 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


