# Learning Topic-Sensitive Word Representations


**Marzieh Fadaee** **Arianna Bisazza** **Christof Monz**

Informatics Institute, University of Amsterdam


## Motivation

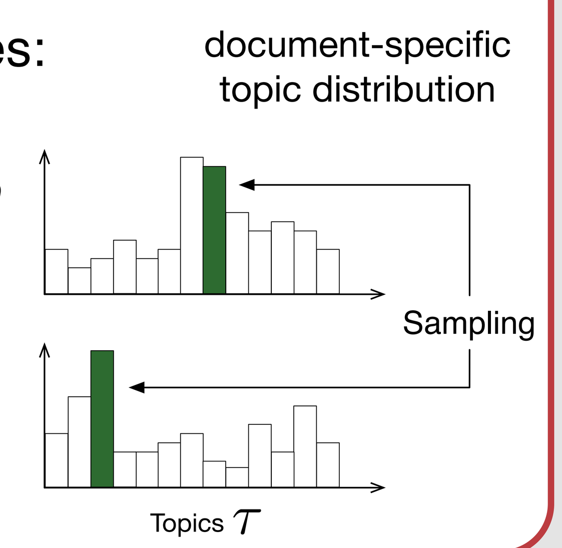
- Having one representation per word fails to capture polysemy
- We propose an approach to learn multiple representations per word by topic-modeling the context with HDP

Polysemous word ↔ Diverse contexts ↔ Distinct topic distributions

## Topic Model: Hierarchical Dirichlet Process (HDP)

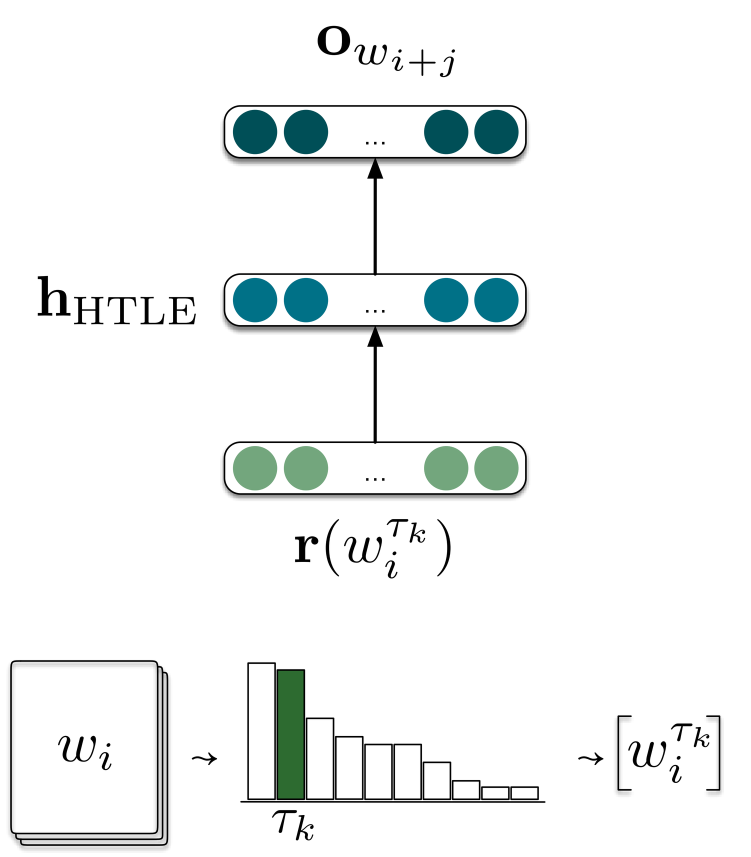
**Example:** The word "bat" in two different sentences:

- *While the team at* ***bat*** *is trying to score runs, the team in the field is attempting to record outs.*
- *The* ***bat*** *wing is a membrane stretched across four "extremely" elongated fingers.*



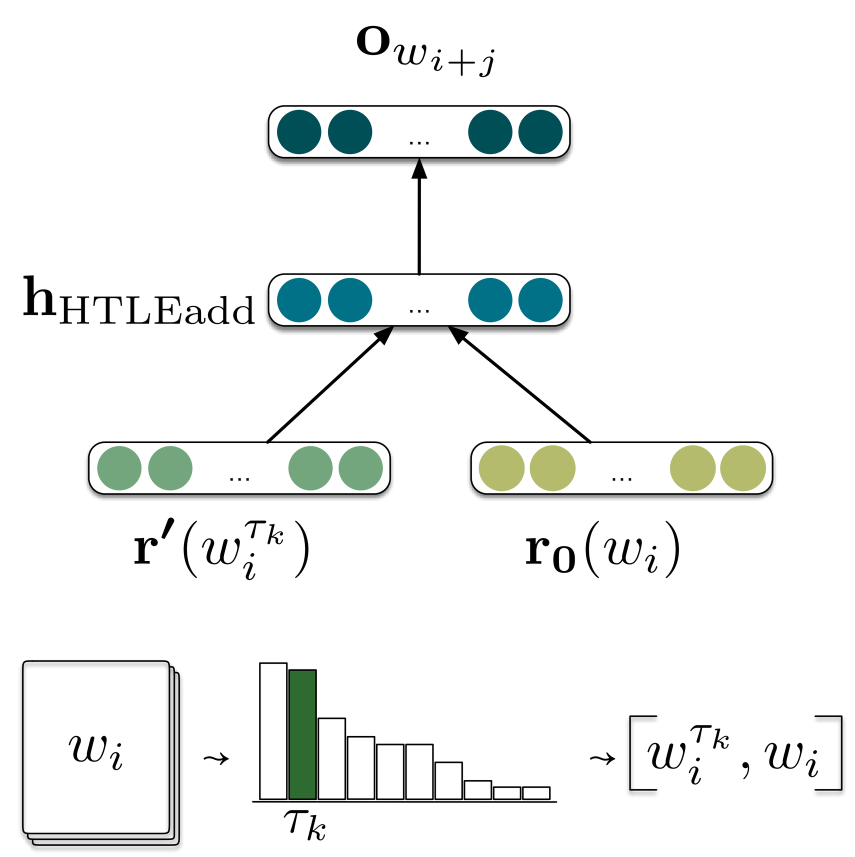

## Topic-Sensitive Representation Models

### Hard Topic-Labeled Representation



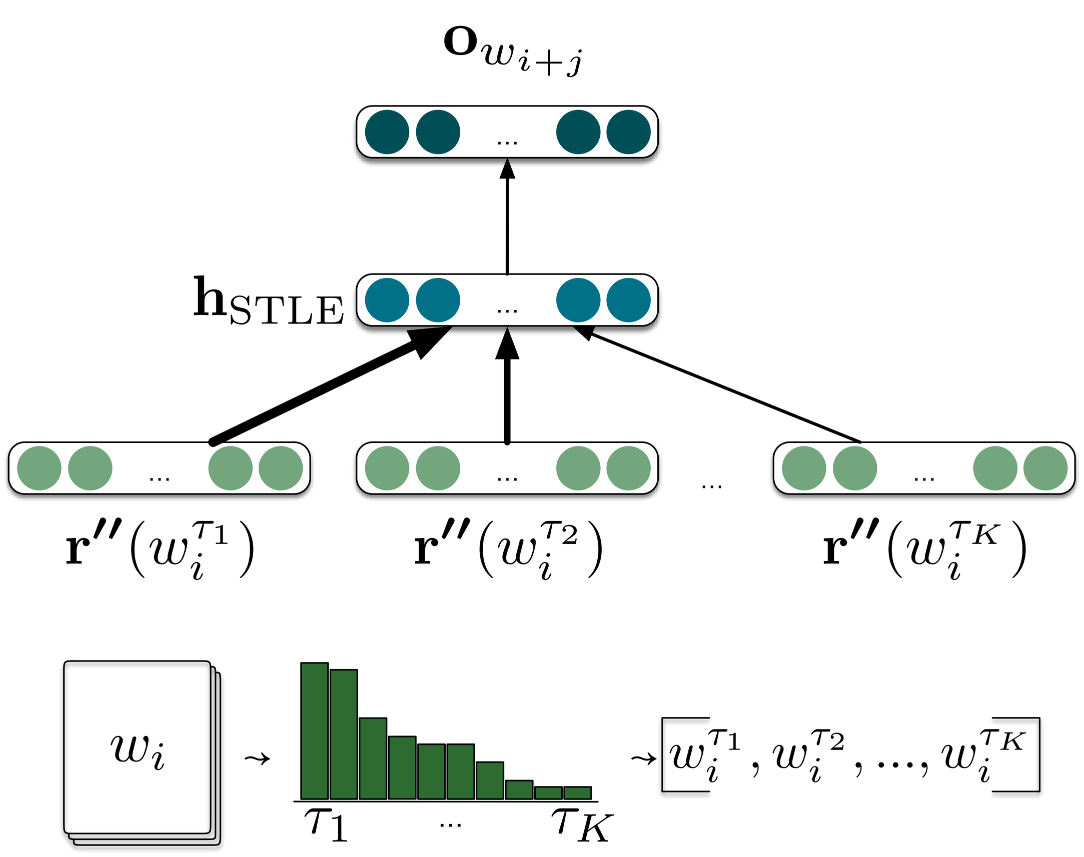

▶ Uses the hard topic labels resulting from HDP sampling to learn representations

### Hard Topic-Labeled + Generic Word Representation




▶ Uses the sum of the hard topic-labeled representation and the generic (i.e. unlabeled) representation

### Soft Topic-Labeled Representation




▶ Uses the topic distribution to compute a weighted sum over the word-topic representations

## Example: Nearest Neighbors of "bat"

| Pre-trained Skipgram | Pre-trained Glove | Skipgram | Topic-Sensitive Skipgram (HTLE) $\tau_1$ | Topic-Sensitive Skipgram (HTLE) $\tau_2$ |
|---|---|---|---|---|
| bats | bats | uroderma | ball | vespertilionidae |
| batting | batting | magnirostrum | pitchout | heran |
| hitter | Bat | sorenseni | batter | hipposideros |
| batsman | catcher | miniopterus | toss-for | sorenseni |
| batted | fielder | prompos | umpire | luctus |
| hoary | hitter | luctus | batting | coxi |
| Batting | balls | micronycteris | fielder | kerivoula |

## Evaluation: Lexical Substitution Task

**Example:** So that in one *way* things in the distressed areas are not as bad as they might be .

- direction
- respect ✓✓
- route
- aspect ✓
- some

$Sim(w_s, w_t)$ → Ranking

1. respect ✓✓
2. aspect ✓
3. direction
4. route
5. some

## Results (Generalized Average Precision)

| Model | Inference | LS-SE07 100 | LS-SE07 300 | LS-SE07 600 | LS-CIC 100 | LS-CIC 300 | LS-CIC 600 |
|---|---|---|---|---|---|---|---|
| SGE + C | N/A | 36.6 | 40.9 | 41.6 | 32.8 | 36.1 | 36.8 |
| MSSG | N/A | 37.8 | 41.1 | 42.9 | 33.9 | 37.8 | 39.1 |
| HTLE | Sampled | 39.8$^{\blacktriangle}$ | 42.5$^{\blacktriangle}$ | 43.0$^{\blacktriangle}$ | 32.1 | 32.7 | 33.0 |
| HTLEadd | Sampled | 39.4$^{\triangle}$ | 41.3$^{\blacktriangle}$ | 41.8 | 30.4 | 31.5 | 31.7 |
| STLE | Sampled | 35.2 | 36.7 | 39.0 | 32.9 | 32.3 | 33.9 |
| HTLE | Expected | **40.3**$^{\blacktriangle}$ | **42.8**$^{\blacktriangle}$ | **43.4**$^{\blacktriangle}$ | 36.6$^{\blacktriangle}$ | **40.9**$^{\blacktriangle}$ | **41.3**$^{\blacktriangle}$ |
| HTLEadd | Expected | 39.9$^{\blacktriangle}$ | 41.8$^{\blacktriangle}$ | 42.2 | 35.5$^{\triangle}$ | 37.9$^{\triangle}$ | 38.6 |
| STLE | Expected | 38.7 | 41.0 | 41.1 | **36.8**$^{\blacktriangle}$ | 36.8 | 37.1 |

$$Sim(w_s, w_t) = cos(\mathbf{h}(w_s^{\tau}), \mathbf{h}(w_t^{\tau'})) + \frac{\sum_c cos(\mathbf{h}(w_s^{\tau}), \mathbf{o}(w_c))}{C}$$

$$Sim(w_s, w_t) = \sum_{\tau,\tau'} p(\tau)\, p(\tau') cos(\mathbf{h}(w_s^{\tau}), \mathbf{h}(w_t^{\tau'})) + \frac{\sum_{\tau,c} cos(\mathbf{h}(w_s^{\tau}), \mathbf{o}(w_c))\; p(\tau)}{C}$$

## Conclusions

- When context is available, multiple representations per word perform best in capturing the underlying meaning
- Our topic-sensitive representations:
  - capture different word senses
  - work as good as Skipgram with 6 times fewer dimensions
  - obtain improvements in the lexical substitution task, performing best in *Noun* substitution


University of Amsterdam



This research was funded in part by the Netherlands Organization for Scientific Research (NWO) under project numbers 639.022.213 and 639.021.646


NWO




m.fadaee@uva.nl