Detecting Asymmetric Semantic Relations in Context

A Case Study on Hypernymy Detection

Yogarshi Vyas and Marine Carpuat

*SEM 2017 - 08/03/2017
Chess \rightarrow \text{Hypernym?} \rightarrow \text{Game}
Magnus Carlsen is the world chess champion. The championship game was played in NYC.
Magnus Carlsen is the world *chess* champion

The championship *game* was played in NYC.
Magnus Carlsen is the world **chess** champion

The poachers hunted the big **game**
Magnus Carlsen is the world *chess* champion

The poachers hunted the big *game*
Magnus Carlsen is the world chess champion.

The championship game was played in NYC.

The poachers hunted the big game.
Hypernymy Detection in Context

Input

Output
Hypernymy Detection in Context

Input

chess

Output

game
Magnus Carlsen is the world chess champion. The championship game was played in NYC.
Hypernymy Detection in Context

Input

Magnus Carlsen is the world **chess** champion

The championship **game** was played in NYC.

Output

✅

❌
Hypernymy Detection in Context

**Input**

Magnus Carlsen is the world **chess** champion

The championship **game** was played in NYC.

**Output**

- ✔️ if game is a hypernym of chess in the given contexts
- ❌
Magnus Carlsen is the world chess champion.

The championship game was played in NYC.

Output:

- ✔️ if game is a hypernym of chess in the given contexts
- ❌ otherwise
Motivation: Why Context?

• Benefit downstream tasks
  - Question Answering
  - Textual Inference
Motivation : Why Context?

• Benefit downstream tasks
  - Question Answering
  - Textual Inference

• Previous work relies on annotator to think about senses
  (Hearst, 1992; Zhitomirsky-Geffet and Dagan, 2009; Turney and Mohammad, 2013)
Previous Work
Previous Work: Context-PPDB

Shwartz and Dagan, *SEM 2016
Previous Work: Context-PPDB

- A dataset for fine-grained lexical inference in context
- 3750 examples, distributed over 6 semantic relation types
Contributions
Contributions

• We construct WHiC, a new dataset of hypernym pairs in context
  • Automatically extracted from WordNet!
Contributions

• We construct **WHiC**, a new dataset of hypernym pairs in context
  
  • Automatically extracted from **WordNet**!

• WHiC enables
  
  • **Empirical comparisons** of context embeddings
Contributions

• We construct WHiC, a new dataset of hypernym pairs in context
  • Automatically extracted from WordNet!

• WHiC enables
  • Empirical comparisons of context embeddings
  • Analysis of sensitivity to context, and directionality
Contributions

- We construct WHiC, a new dataset of hypernym pairs in context
  - Automatically extracted from WordNet!

- WHiC enables
  - **Empirical comparisons** of context embeddings
  - Analysis of **sensitivity to context**, and **directionality**

- We provide experimental evidence of **challenging** nature of task
Outline

• WHiC : A Dataset for Hypernymy Detection in Context
  • Dataset Desiderata
  • Dataset Creation

• Two Models of Context Representation

• Experiments and Analysis
Outline

• WHiC : A Dataset for Hypernymy Detection in Context
  • Dataset Desiderata
  • Dataset Creation

• Two Models of Context Representation

• Experiments and Analysis
Dataset Desiderata
Dataset Desiderata

• Test sensitivity of models to changing contexts
Dataset Desiderata

- Test sensitivity of models to changing contexts
- Differentiate between hypernymy and semantic similarity
  - (Chess, Game) \textit{v/s} (Chess, Checkers)
Detecting Asymmetric Semantic Relations in Context

Yogarshi Vyas

Dataset Desiderata

• Test sensitivity of models to changing contexts

• Differentiate between hypernymy and semantic similarity
  • (Chess, Game) v/s (Chess, Checkers)

• Large enough to train supervised models
WHiC : WordNet Hypernyms in Context
WHiC : WordNet Hypernyms in Context

- Word: Representation
  Example: the drawing was not a faithful representation

- Word: Drawing, Art
  Example: he did complicated pen-and-ink drawings

- Word: Study, Sketch
  Example: he made several studies before the final painting
WHiC: WordNet Hypernyms in Context

- Word: Representation
  Example: the drawing was not a faithful representation

- Word: Drawing, Art
  Example: he did complicated pen-and-ink drawings

- Word: Study, Sketch
  Example: he made several studies before the final painting
WHiC : WordNet Hypernyms in Context

Word: Study, Sketch
Example: he made several studies before the final painting

Word: Drawing, Art
Example: he did complicated pen-and-ink drawings

Word: Representation
Example: the drawing was not a faithful representation
WHiC : Positive Examples

Word : Study
Example : he knocked on the door of the study

Word : Study
Example : he made several studies before the final painting
WHiC : Positive Examples

**Word**: Room  
**Example**: the rooms were small but comfortable

**Word**: Study  
**Example**: he knocked on the door of the study

**Word**: Drawing  
**Example**: he did complicated pen-and-ink drawings

**Word**: Study  
**Example**: he made several studies before the final painting

\[
\text{IsHypernym}(w_1, w_2) ?
\]
WHiC: Positive Examples

Word: Room
Example: the rooms were small but comfortable

Word: Study
Example: he knocked on the door of the study

Word: Drawing
Example: he did complicated pen-and-ink drawings

Word: Study
Example: he made several studies before the final painting

<table>
<thead>
<tr>
<th>$w_1$, $w_2$</th>
<th>$C_1$, $C_2$</th>
<th>$\text{IsHypernym}(w_1, w_2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study, Room</td>
<td>He knocked on the door of the study</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>The rooms were small but comfortable</td>
<td></td>
</tr>
<tr>
<td>Study, Drawing</td>
<td>He made 2 studies before the painting</td>
<td>✔</td>
</tr>
<tr>
<td></td>
<td>He did pen-and-ink drawings</td>
<td></td>
</tr>
</tbody>
</table>
WHiC: Negative Examples

<table>
<thead>
<tr>
<th>Word</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Room</td>
<td>The rooms were small but comfortable</td>
</tr>
<tr>
<td>Study</td>
<td>He knocked on the door of the study</td>
</tr>
<tr>
<td>Drawing</td>
<td>He did complicated pen-and-ink drawings</td>
</tr>
<tr>
<td>Study</td>
<td>He made several studies before the final painting</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$w_1, w_2$</th>
<th>$c_1, c_2$</th>
<th>$\text{IsHypernym}(w_1, w_2)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study, Room</td>
<td>He knocked on the door of the study, The rooms were small but comfortable</td>
<td>✓</td>
</tr>
<tr>
<td>Study, Drawing</td>
<td>He made 2 studies before the painting, He did pen-and-ink drawings</td>
<td>✓</td>
</tr>
</tbody>
</table>
### WHiC : Negative Examples

<table>
<thead>
<tr>
<th>Word</th>
<th>Example</th>
<th>( \text{IsHypernym}(w_1, w_2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study, Room</td>
<td>He knocked on the door of the <strong>study</strong></td>
<td>✔️</td>
</tr>
<tr>
<td></td>
<td>The <strong>rooms</strong> were small but comfortable</td>
<td></td>
</tr>
<tr>
<td>Study, Drawing</td>
<td>He made 2 <strong>studies</strong> before the painting</td>
<td>✔️</td>
</tr>
<tr>
<td></td>
<td>He did pen-and-ink <strong>drawings</strong></td>
<td></td>
</tr>
<tr>
<td>Study, Room</td>
<td>He knocked on the door of the <strong>study</strong></td>
<td>✗</td>
</tr>
<tr>
<td></td>
<td>He did pen-and-ink <strong>drawings</strong></td>
<td></td>
</tr>
</tbody>
</table>
### WHiC: Negative Examples

<table>
<thead>
<tr>
<th>Word</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study</td>
<td>he knocked on the door of the study</td>
</tr>
<tr>
<td>Room</td>
<td>The rooms were small but comfortable</td>
</tr>
<tr>
<td>Study</td>
<td>he made several studies before the final painting</td>
</tr>
<tr>
<td>Drawing</td>
<td>he did complicated pen-and-ink drawings</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$w_1, w_2$</th>
<th>$c_1, c_2$</th>
<th>IsHypernym$(w_1, w_2)$?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study, Room</td>
<td>He knocked on the door of the study</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>The rooms were small but comfortable</td>
<td></td>
</tr>
<tr>
<td>Study, Drawing</td>
<td>He made 2 studies before the painting</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>He did pen-and-ink drawings</td>
<td></td>
</tr>
<tr>
<td>Study, Room</td>
<td>He knocked on the door of the study</td>
<td>✗</td>
</tr>
<tr>
<td></td>
<td>He did pen-and-ink drawings</td>
<td></td>
</tr>
</tbody>
</table>
## WHiC : Negative Examples

<table>
<thead>
<tr>
<th>Word</th>
<th>Example</th>
<th>Word</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Room</td>
<td>the rooms were small but comfortable</td>
<td>Drawing</td>
<td>he did complicated pen-and-ink drawings</td>
</tr>
<tr>
<td>Study</td>
<td>he knocked on the door of the study</td>
<td>Study</td>
<td>he made several studies before the final painting</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$w_1, w_2$</th>
<th>$C_1, C_2$</th>
<th>IsHypernym($w_1, w_2$) ?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study, Room</td>
<td>He knocked on the door of the study</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>The rooms were small but comfortable</td>
<td>✓</td>
</tr>
<tr>
<td>Study, Drawing</td>
<td>He made 2 studies before the painting</td>
<td>✓</td>
</tr>
<tr>
<td></td>
<td>He did pen-and-ink drawings</td>
<td>✓</td>
</tr>
<tr>
<td>Study, Room</td>
<td>He knocked on the door of the study</td>
<td>✗</td>
</tr>
<tr>
<td></td>
<td>He did pen-and-ink drawings</td>
<td>✗</td>
</tr>
<tr>
<td>Room, Study</td>
<td>The rooms were small but comfortable</td>
<td>✗</td>
</tr>
<tr>
<td></td>
<td>He knocked on the door of the study</td>
<td>✗</td>
</tr>
</tbody>
</table>
Dataset Desiderata

• Test sensitivity of models to changing contexts ✔

• Differentiate between **hypernymy** and **semantic similarity** ✔

• Large enough to train supervised learning models on ✔
  - Seed vocab = 10k most frequent words from Wikipedia
  - 22000 examples distributed over 6000 word pairs
Outline

- WHiC: A Dataset for Hypernymy Detection in Context
  - Dataset Desiderata
  - Dataset Creation
- Two Models of Context Representation
- Experiments and Analysis
Outline

- WHiC : A Dataset for Hypernymy Detection in Context
  - Dataset Desiderata
  - Dataset Creation
- Two Models of Context Representation
- Experiments and Analysis
Supervised Hypernymy Detection

chess

game
Supervised Hypernymy Detection

chess

game
Supervised Hypernymy Detection

\[ f(\text{chess}, \text{game}) \]

Baroni et al., 2012; Roller et al., 2014; Weeds et al., 2014 *inter alia*
Supervised Hypernymy Detection

\[ f(\text{chess}, \text{game}) \]

\[ x \cdot y, y - x, [x; y] \]

Baroni et al., 2012; Roller et al., 2014; Weeds et al., 2014 *inter alia*
Hypernymy-Feature Detector

Roller and Erk; EMNLP 2016
Game

Word (w)
Word (w)

Game

The hunters hunted the game

Contextualized Word (w_c)
<table>
<thead>
<tr>
<th>Word (w)</th>
<th>Context (c)</th>
<th>Contextualized Word (w_c)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>The hunters hunted the game</td>
<td>The hunters hunted the <strong>game</strong></td>
</tr>
<tr>
<td></td>
<td>Game</td>
<td></td>
</tr>
</tbody>
</table>
Detecting Asymmetric Semantic Relations in Context

X

=}

Erk and Padó, EMNLP 2008; Thater et al., IJCNLP 2011; Dinu et al., NAACL 2012
The hunters hunted the game

The hunters hunted the game

Detecting Asymmetric Semantic Relations in Context
Representing Context

Two complementary approaches
Representing Context

Two complementary approaches

- **Pooling**: Combining word type representations into context representations
  - Mean, min, and max
Representing Context : Pooling
Representing Context: Pooling

<table>
<thead>
<tr>
<th>the</th>
<th>0.3</th>
<th>0.6</th>
<th>-0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>river</td>
<td>1.5</td>
<td>-2.5</td>
<td>0</td>
</tr>
<tr>
<td>bank</td>
<td>-1</td>
<td>0.2</td>
<td>1.8</td>
</tr>
</tbody>
</table>
Representing Context: Pooling

<table>
<thead>
<tr>
<th></th>
<th>the</th>
<th>river</th>
<th>bank</th>
</tr>
</thead>
<tbody>
<tr>
<td>the</td>
<td>0.3</td>
<td>0.6</td>
<td>-0.1</td>
</tr>
<tr>
<td>river</td>
<td>1.5</td>
<td>-2.5</td>
<td>0</td>
</tr>
<tr>
<td>bank</td>
<td>-1</td>
<td>0.2</td>
<td>1.8</td>
</tr>
</tbody>
</table>

\[ \overrightarrow{C}_{l, \text{mean}} = \begin{bmatrix} 0.27 \\ -0.57 \\ 0.57 \end{bmatrix} \]
Representing Context: Pooling

<table>
<thead>
<tr>
<th></th>
<th>the</th>
<th>river</th>
<th>bank</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.3</td>
<td>0.6</td>
<td>-0.1</td>
</tr>
<tr>
<td></td>
<td>1.5</td>
<td>-2.5</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>-1</td>
<td>0.2</td>
<td>1.8</td>
</tr>
</tbody>
</table>

$\vec{C}_{l,\text{mean}}$: 0.27 -0.57 0.57

$\vec{C}_{l,\text{min}}$: -1 -2.5 -0.1

Yogarshi Vyas
Representing Context: Pooling

<table>
<thead>
<tr>
<th>the</th>
<th>0.3</th>
<th>0.6</th>
<th>-0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>river</td>
<td>1.5</td>
<td>-2.5</td>
<td>0</td>
</tr>
<tr>
<td>bank</td>
<td>-1</td>
<td>0.2</td>
<td>1.8</td>
</tr>
</tbody>
</table>

$\vec{C}_{l,\text{mean}}$

<table>
<thead>
<tr>
<th></th>
<th>0.27</th>
<th>-0.57</th>
<th>0.57</th>
</tr>
</thead>
</table>

$\vec{C}_{l,\text{min}}$

<table>
<thead>
<tr>
<th></th>
<th>-1</th>
<th>-2.5</th>
<th>-0.1</th>
</tr>
</thead>
</table>

$\vec{C}_{l,\text{max}}$

|          | 1.5  | 0.6   | 1.8  |
# Representing Context: Pooling

<table>
<thead>
<tr>
<th></th>
<th>the</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.3</td>
<td>0.6</td>
<td>-0.1</td>
</tr>
<tr>
<td>river</td>
<td>1.5</td>
<td>-2.5</td>
<td>0</td>
</tr>
<tr>
<td>bank</td>
<td>-1</td>
<td>0.2</td>
<td>1.8</td>
</tr>
</tbody>
</table>

\[ \vec{C}_{l,\text{mean}} = \begin{bmatrix} 0.27 \\ -0.57 \\ 0.57 \end{bmatrix} \]

\[ \vec{C}_{l,\text{min}} = \begin{bmatrix} -1 \\ -2.5 \\ -0.1 \end{bmatrix} \]

\[ \vec{C}_{l,\text{max}} = \begin{bmatrix} 1.5 \\ 0.6 \\ 1.8 \end{bmatrix} \]

Tang et al., ACL 2014; Hovy, ACL 2015
Representing Context: Pooling

\[ \vec{c}_l = [\vec{c}_{l,\text{mean}} ; \vec{c}_{l,\text{min}} ; \vec{c}_{l,\text{max}}] \]

<table>
<thead>
<tr>
<th>the</th>
<th>0.3</th>
<th>0.6</th>
<th>-0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>river</td>
<td>1.5</td>
<td>-2.5</td>
<td>0</td>
</tr>
<tr>
<td>bank</td>
<td>-1</td>
<td>0.2</td>
<td>1.8</td>
</tr>
</tbody>
</table>

\[ \vec{c}_{l,\text{mean}} = [0.27, -0.57, 0.57] \]
\[ \vec{c}_{l,\text{min}} = [-1, -2.5, -0.1] \]
\[ \vec{c}_{l,\text{max}} = [1.5, 0.6, 1.8] \]

Tang et al., ACL 2014; Hovy, ACL 2015
Representing Context

Two complementary approaches

- **Pooling**: Combining word type representations into context representations
  - Mean, min, and max
Representing Context

Two complementary approaches

- **Pooling**: Combining word type representations into context representations
  - Mean, min, and max

- **RNN Composition**: Embed words and contexts jointly
Representing Context: Context2Vec

Melamud et al; CoNLL 2016
Representing Context : Context2Vec

Image courtesy : context2vec: Learning Generic Context Embedding with Bidirectional LSTM, Melamud et al., CoNLL 2016
Detecting Asymmetric Semantic Relations in Context

Word (w)  X  Context (c)  =  Contextualized Word (w_c)
Detecting Asymmetric Semantic Relations in Context

\[ \text{Word (w)} \times \text{Context (c)} = \text{Contextualized Word (w_c)} \]
Putting it all together ..

- **Task**: Context Aware Hypernymy Detection
Putting it all together ..

- **Task**: Context Aware Hypernymy Detection
- **Dataset**: WHiC
Putting it all together ..

- **Task**: Context Aware Hypernymy Detection
- **Dataset**: WHiC
- **Classifier**: H-Feature Detector
Putting it all together ..

- **Task**: Context Aware Hypernymy Detection
- **Dataset**: WHiC
- **Classifier**: H-Feature Detector
- **Input to classifier**: Context aware word representations
  - Pooling
  - Context2Vec
Outline

- WHiC: A Dataset for Hypernymy Detection in Context
  - Dataset Desiderata
  - Dataset Creation
- Two Models of Context Representation
- Experiments and Analysis
Overall Results
Detecting Asymmetric Semantic Relations in Context

Yogarshi Vyas

F-Score on WHiC

- **Words (w)**
- **Context-aware words (wc)**
- **w + wc**

Pooling

C2V

Hybrid

0 10 20 30 40 50 60 70 80
Detecting Asymmetric Semantic Relations in Context

Yogarshi Vyas

Pooling

C2V

Hybrid

Words (w)  Context-aware words (wc)  w + wc

F-Score on WHiC

Words (w)  Context-aware words (wc)  w + wc

Pooling

C2V

Hybrid

0 10 20 30 40 50 60 70 80

F-Score on WHiC
Detecting Asymmetric Semantic Relations in Context

Yogarshi Vyas

- **Pooling**
- **C2V**
- **Hybrid**

The bar chart illustrates the F-Score on WHiC for different methods. The x-axis represents the F-Score ranging from 0 to 80, while the y-axis lists the methods: **Pooling**, **C2V**, and **Hybrid**. Each method is represented by three bars, indicating the performance of **Words (w)**, **Context-aware words (wc)**, and **w + wc**. The chart shows a comparison of these methods in detecting asymmetric semantic relations in context.
Detecting Asymmetric Semantic Relations in Context

- **Pooling**
  - Words (w)
  - Context-aware words (wc)
  - \( w + wc \)

- **C2V**
  - Words (w)
  - Context-aware words (wc)
  - \( w + wc \)

- **Hybrid**
  - Words (w)
  - Context-aware words (wc)
  - \( w + wc \)

- **F-Score on WHiC**

- **All models > “All True” Baseline**
• All models > “All True” Baseline
• Context aware representations help ..
• All models > “All True” Baseline
• Context aware representations help ..
• All models > “All True” Baseline
• Context aware representations help ..
• Hybrid model works best
All models > “All True” Baseline

Context aware representations help ..

Hybrid model works best

Small improvements over context agnostic models
Analysis: Sensitivity to Context
Analysis : Sensitivity to Context

**Word** : Room  
**Example** : the rooms were small but comfortable

**Word** : Study  
**Example** : he knocked on the door of the study

**Word** : Drawing  
**Example** : he did complicated pen-and-ink drawings

**Word** : Study  
**Example** : he made several studies before the final painting
Analysis: Sensitivity to Context

Word: Room
Example: the rooms were small but comfortable

Word: Study
Example: he knocked on the door of the study

Word: Drawing
Example: he did complicated pen-and-ink drawings

Word: Study
Example: he made several studies before the final painting
Detecting Asymmetric Semantic Relations in Context

Yogarshi Vyas

- **Pooling**
- **C2V**
- **Hybrid**

Words (w) vs. Context-aware words (wc) vs. w + wc

Context Sensitivity F-Score

0 10 20 30 40 50 60 70 80
• Context aware representations do better on this subset
- Context aware representations do better on this subset..
- .. followed by the hybrid model
Analysis : Directionality
Analysis: Directionality

**Word**: Room  
**Example**: the rooms were small but comfortable

**Word**: Study  
**Example**: he knocked on the door of the study

**Word**: Drawing  
**Example**: he did complicated pen-and-ink drawings

**Word**: Study  
**Example**: he made several studies before the final painting
Analysis: Directionality

Word: Room
Example: the rooms were small but comfortable

Word: Study
Example: he knocked on the door of the study

Word: Drawing
Example: he did complicated pen-and-ink drawings

Word: Study
Example: he made several studies before the final painting
Detecting Asymmetric Semantic Relations in Context

Yogarshi Vyas

Pooling

C2V

Hybrid

Directionality Score

Words (w)  Context-aware words (wc)  w + wc

0  10  20  30  40  50  60  70  80

Detecting Asymmetric Semantic Relations in Context
Detecting Asymmetric Semantic Relations in Context

Yogarshi Vyas

- Context-aware representations do worse
• Context-aware representations do worse
• Context2Vec models are better than Pooling models
• Context-aware representations do worse
• Context2Vec models are better than Pooling models
- Context-aware representations do worse
- Context2Vec models are better than Pooling models
- Hybrid model is best - mirroring performance on the full dataset
Results Summary

• Context aware hypernymy detection is challenging!
Results Summary

- Context aware hypernymy detection is challenging!

- Context aware representations help in capturing context
  - Do worse at capturing asymmetry
Results Summary

• Context aware hypernymy detection is **challenging**!

• Context aware representations **help** in capturing **context**
  • Do **worse** at capturing **asymmetry**

• **Hybrid models** can combine strengths of multiple models
Contributions

• We construct WHiC, a new dataset of hypernym pairs in context
  • Automatically extracted from WordNet!
Contributions

• We construct **WHiC**, a new dataset of hypernym pairs in context
  - Automatically extracted from **WordNet**!

• WHiC enables
  - *Empirical comparisons* of context embeddings
  - Analysis of *sensitivity to context*, and *directionality*
Contributions

• We construct **WHiC**, a new dataset of hypernym pairs in context
  - Automatically extracted from **WordNet**!

• WHiC enables
  - **Empirical comparisons** of context embeddings
  - Analysis of **sensitivity to context**, and **directionality**

• We provide experimental evidence of **challenging** nature of task
Contributions

• We construct WHiC, a new dataset of hypernym pairs in context
  - Automatically extracted from WordNet!

• WHiC enables
  - Empirical comparisons of context embeddings
  - Analysis of sensitivity to context, and directionality

• We provide experimental evidence of challenging nature of task

https://github.com/yogarshi/WHiC
Backup
Supervised Hypernymy Detection

\( f(\text{chess} - \text{game}) \)
Supervised Hypernymy Detection

\[ f(\text{chess}, \text{game}) \]
Supervised Hypernymy Detection

\( f(\text{chess}, \text{game}) \)
Supervised Hypernymy Detection

\[ f(\text{dog}, \text{game}) \]
Supervised Hypernymy Detection
Supervised Hypernymy Detection

\[ f(\text{sofa}, \text{!!!!}, \text{game}) \]
H - Feature Detector

Roller and Erk; EMNLP 2016
H - Feature Detector

• Learn a separating hyperplane using a linear classifier

• Project word representations on the plane orthogonal to this hyperplane
H - Feature Detector

- Learn a separating hyperplane using a linear classifier
- Project word representations on the plane **orthogonal** to this hyperplane

Repeat
H - Feature Detector

• Learn a separating hyperplane using a linear classifier

• Project word representations on the plane orthogonal to this hyperplane

Repeat

Iteration $i$ — words $x_i$ and $y_i$ — learned hyperplane $h_i$
H - Feature Detector

- Learn a separating hyperplane using a linear classifier
- Project word representations on the plane \textit{orthogonal} to this hyperplane

\small

Iteration $i$ — words $x_i$ and $y_i$ — learned hyperplane $h_i$

1. $x_i \cdot y_i$
**H - Feature Detector**

- Learn a separating hyperplane using a linear classifier
- Project word representations on the plane **orthogonal** to this hyperplane

---

Iteration $i$ — words $\mathbf{x}_i$ and $\mathbf{y}_i$ — learned hyperplane $\mathbf{h}_i$

1. $\mathbf{x}_i \cdot \mathbf{y}_i$
2. $\mathbf{x}_i \cdot \mathbf{h}_i$
H - Feature Detector

• Learn a separating hyperplane using a linear classifier

• Project word representations on the plane orthogonal to this hyperplane

Repeat

Iteration $i$ — words $x_i$ and $y_i$ — learned hyperplane $h_i$

1. $x_i \cdot y_i$
2. $x_i \cdot h_i$
3. $y_i \cdot h_i$
\[ H - \text{Feature Detector} \]

- Learn a separating hyperplane using a linear classifier
- Project word representations on the plane **orthogonal** to this hyperplane

**Repeat**

**Iteration** \( i \) — words \( x_i \) and \( y_i \) — learned hyperplane \( h_i \)

1. \( x_i \cdot y_i \)
2. \( x_i \cdot h_i \)
3. \( y_i \cdot h_i \)
4. \( (y_i - x_i) \cdot h_i \)
H - Feature Detector

- Learn a separating hyperplane using a linear classifier
- Project word representations on the plane **orthogonal** to this hyperplane

Repeat

**Iteration** $i$ — words $x_i$ and $y_i$ — learned hyperplane $h_i$

1. $x_i \cdot y_i$
2. $x_i \cdot h_i$
3. $y_i \cdot h_i$
4. $(y_i - x_i) \cdot h_i$