A Two Level Model for Context Sensitive Inference Rules

Oren Melamud, Jonathan Berant, Ido Dagan, Jacob Goldberger and Idan Szpektor
Context sensitive inference

\[
treat \rightarrow kill
\]

- medicine \textit{treat} bacteria
- doctor \textit{treat} patient
- medicine \textit{kill} bacteria
- doctor \textit{kill} patient

\checkmark \quad \times
Outline

1. Context Sensitive Inference
2. Two Level Model
3. Results
context sensitive inference

two level model

results
Textual inference

Text (t)

Today, Google Inc. acquired Waze Ltd. for over a billion dollars.

Hypothesis (h)

Google purchased a startup

Useful in various NLP applications (QA, IE, etc.)
Textual inference rules

Text (t)

Today, Google Inc. acquired Waze Ltd. for over a billion dollars.

Hypothesis (h)

Google purchased a startup

Inference rules

category: context sensitive inference
two level model
results
Today, Google Inc. acquired Waze Ltd. for over a billion dollars.

Google purchased a startup
Predicate templates

X slot

acquire

Y slot

context sensitive inference

two level model

results

Bar Ilan University @ ACL 2013
Predicate templates

Google acquire Waze

X slot acquire Y slot

context sensitive inference two level model results
Predicate inference rules
(not context sensitive)

<table>
<thead>
<tr>
<th>Rule</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>X acquire Y ↔ X purchase Y</td>
<td>0.5</td>
</tr>
<tr>
<td>X like Y → X know Y</td>
<td>0.7</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rule application</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>John likes Mary → John knows Mary</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Existing resources:
DIRT (Lin and Pantel 2001), Berant et al. 2011, etc.
But inference is context sensitive

\[ X \text{ treat } Y \rightarrow X \text{ kill } Y \]

- medicine \textit{treat} bacteria
  - high score
  - medicine \textit{kill} bacteria
- doctor \textit{treat} patient
  - low score
  - doctor \textit{kill} patient

context sensitive inference  \rightarrow  two level model  \rightarrow  results
Addressing context sensitive inference

Previous work:

- Several lines of work
- Predicate inference:
  - Semantic classes
  - Latent topic models (LDA)
context sensitive inference

two level model

results
Outline

Word-level context insensitive models (DIRT-like)

Topic-level context sensitive models (LDA)

Prior work

Our work

Two-level word and topic model

countext sensitive inference ➔ two level model ➔ results
### Outline

- **Word-level context insensitive models (DIRT-like)**
- **Topic-level context sensitive models (LDA)**

#### Prior work

#### Our work

- Two-level word and topic model

**Steps:**
- Context sensitive inference
- Two level model
- Results
Outline

Word-level context insensitive models (DIRT-like)

Topic-level context sensitive models (LDA)

Prior work

Our work

Two-level word and topic model
Outline

Word-level context insensitive models (DIRT-like)

Topic-level context sensitive models (LDA)

Prior work

Our work

Two-level word and topic model
### Context insensitive rules (DIRT-like)

<table>
<thead>
<tr>
<th>u</th>
<th>acquire Y</th>
<th>v</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>0.5</td>
<td>company</td>
<td>0.3</td>
</tr>
<tr>
<td>1.2</td>
<td>Waze</td>
<td>1.5</td>
</tr>
<tr>
<td>0.7</td>
<td>Skype</td>
<td>1.2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>0.5</td>
<td>skills</td>
<td>...</td>
</tr>
<tr>
<td>0.2</td>
<td>data</td>
<td>0.1</td>
</tr>
<tr>
<td>0.3</td>
<td>language</td>
<td>...</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

$$v[w] = PMI(pred, w)$$
Word level “base” models

Context insensitive rules (DIRT-like)

<table>
<thead>
<tr>
<th>u</th>
<th>acquire Y</th>
<th>v</th>
</tr>
</thead>
<tbody>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>0.5</td>
<td>company</td>
<td>0.3</td>
</tr>
<tr>
<td>1.2</td>
<td>Waze</td>
<td>1.5</td>
</tr>
<tr>
<td>0.7</td>
<td>Skype</td>
<td>1.2</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>0.5</td>
<td>skills</td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>data</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>language</td>
<td></td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

**sim measure example:**

\[
score(u, v) = \sum_w u[w] \cdot v[w]
\]
Topic level models

Context sensitive rules
Dinu and Lapata 2010, Ritter et al. 2010

acquire Y

...  
company  
Waze  
Skype  
...  
skills  
data  
language  
...

purchase Y

...  
company  
Waze  
Skype  
...  
-  
data  
-  
...

two level model

context sensitive inference

results

Bar Ilan University @ ACL 2013 20
Topic level models

Step 1: Identify latent dimensions

acquire Y

company, Waze, Skype, ...
skills, data, language

pseudo-documents

two level model

purchase Y

company, Waze, Skype, ...
data

context sensitive inference

results

Bar Ilan University @ ACL 2013
Topic level models

Step 1: Identify latent dimensions

acquire Y
company, Waze, Skype, ...

purchase Y
company, Waze, Skype, ...

LDA topics

skills, data, language

data

context sensitive inference ➔ two level model ➔ results

Bar Ilan University @ ACL 2013
Step 1: Identify latent dimensions

LDA topics

company, Waze, Skype, ...

skills, data, language

data

context sensitive inference
two level model
results

p(t|pred)

acquire Y

company, Waze, Skype, ...

purchase Y

company, Waze, Skype, ...

Topic level models

Bar Ilan University @ ACL 2013
Topic level models

Step 2: Topic vectors replace word vectors

acquire Y
- ...
- company
- Waze
- Skype
- ...
- skills
- data
- language
- ...

purchase Y
- ...
- company
- Waze
- Skype
- ...
- -
- data
- -
- ...

context sensitive inference ➔ two level model ➔ results

Bar Ilan University @ ACL 2013
Topic level models

Step 2: Topic vectors replace word vectors

acquire \( Y \)

\[ p(t \mid \text{pred}) \]

purchase \( Y \)

\[ p(t \mid \text{pred}) \]
Step 3: Context sensitive topic vectors and scores

**acquire Y**

\[ p(t | \text{pred}) \]

**acquire IBM**

\[ p(t | \text{pred}, w) \]

**purchase Y**

\[ p(t | \text{pred}) \]

**context sensitive inference**

**two level model**

**results**

\[ p(t | \text{pred}, w) = \frac{1}{Z} p(t | \text{pred}) \cdot p(w | t) \]
Topic level models

Step 3: Context sensitive topic vectors and scores

- acquire $Y$
  - $p(t|\text{pred})$

- acquire skill
  - $p(t|\text{pred},w)$

- purchase $Y$
  - $p(t|\text{pred})$

context sensitive inference  ➔  two level model  ➔  results
The best of all worlds

Prior work

Our work

Word-level
+ Fine grained
- Context insensitive

Topic-level
- Coarse grained
+ Context sensitive

Two-level
+ Fine grained
+ Context sensitive

context sensitive inference ➔ two level model ➔ results
Our contribution: Two level model

Step 1: Identify latent dimensions

acquire Y

- company
- Waze
- Skype
- ...

purchase Y

- company
- Waze
- Skype
- ...

skills

data

language

- ...

context sensitive inference  two level model  results

Bar Ilan University @ ACL 2013
Two level model

Step 2: Learn inference score per topic

acquire Y

...  
company  
Waze  
Skype  
...  

purchase Y

...  
company  
Waze  
Skype  
...  

skills  
data  
language  
...  

context sensitive inference  
two level model  
results
Two level model

Step 2: Learn inference score per topic

\[ \text{score}_{\text{red}}(u, v) = 0.9 \]

\[ v_t[w] = v[w] \cdot p(t|\text{pred}, w) \]

\[ \text{score}_t(u, v) = \text{score}(u_t, v_t) \]
Two level model

Step 2: Learn inference score per topic

- acquire Y
  - company
  - Waze
  - Skype
  - skills
  - data
  - language
  - ...

- purchase Y
  - company
  - Waze
  - Skype
  - data
  - ...

score per each topic replaces traditional single score

context sensitive inference  two level model  results

Bar Ilan University @ ACL 2013
Two level model

Step 3: Rule application score = weighted average of per-topic scores

acquire Y

- company
- Waze
- Skype

purchase Y

- company
- Waze
- Skype

skills

- data

language

0.9

0.1

context sensitive inference  two level model  results

Bar Ilan University @ ACL 2013
Two level model

Step 3: Rule application score = weighted average of per-topic scores

acquire IBM

0.8

0.2

acquire Y

company

Waze

Skype

purchase Y

company

Waze

Skype

skills

data

language

context sensitive inference

two level model

results

Bar Ilan University @ ACL 2013

0.8 x 0.9 = 0.74
Two level model

Step 3: Rule application score = weighted average of per-topic scores

acquire skill

p(t|pred,w)

0.2

0.8

acquire Y

company

Waze

Skype

Waze

Skype

purchase Y

company

skills

data

language

Context sensitive inference → Two level model → Results

Bar Ilan University @ ACL 2013
Summary of contribution

Previous approach

- acquire IBM
- purchase Y
- topic vector

Our hybrid approach

- acquire IBM
- acquire Y
- purchase Y
- topic vector
- word vector

context sensitive inference  two level model  results
context sensitive inference

two level model

results
Evaluation settings

**Learning Corpus**
- ReVerb (Fader et al. 2011)
- 15M *extractions* from the web

**Test set**
- Zeichner et al. 2012
- 6,500 *rule applications*
- We used subsets of this dataset
Evaluation settings

Learning Corpus
- ReVerb (Fader et al. 2011)
- 15M extractions from the web

Test set
- Zeichner et al. 2012
- 6,500 rule applications
- We used subsets of this dataset

Test set is not biased towards ambiguous predicates
## Compared models

<table>
<thead>
<tr>
<th>Model</th>
<th>Rule application score</th>
<th>Optimal #topics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Insensitive (&quot;base&quot;)</td>
<td>DIRT (word level model)</td>
<td>NA</td>
</tr>
<tr>
<td>Sensitive I</td>
<td>Ritter et al. 2010 similarity (topic level)</td>
<td>1200</td>
</tr>
<tr>
<td>Sensitive II</td>
<td>Dinu &amp; Lapata 2010 similarity (topic level)</td>
<td>1200</td>
</tr>
<tr>
<td>Our model</td>
<td>Two level model</td>
<td>100</td>
</tr>
</tbody>
</table>

Sensitive baselines require much more topics as they capture all the information in the topic vectors.

All methods are unsupervised and use the same learning corpus.
Results

Similar results for different word-level similarity measures (Lin, Cosine, BInc).

**Rule applications MAP values**

<table>
<thead>
<tr>
<th>Mean Average Precision</th>
<th>Insensitive (&quot;base&quot;)</th>
<th>Sensitive I</th>
<th>Sensitive II</th>
<th>Our model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.503</td>
<td>0.451</td>
<td>0.443</td>
<td>0.562</td>
</tr>
</tbody>
</table>

context sensitive inference ➤ two level model ➤ results

Bar Ilan University @ ACL 2013
Analysis

Effect of test set’s context sensitivity bias

• Context sensitive models perform relatively worse on test sets that are not biased to context sensitivity.
• Our model performs well even on a “context insensitive” test set
Example: Top-scored inferences

- baby acquire skill
  - Our model
  - Insensitive “base” model

- baby develop skill
- baby buy skill
Example: Top-scored inferences

- aspirin **fight** headache
- aspirin **help** headache
- aspirin **lose** headache

- **Our model**
- **Insensitive “base” model**

- context sensitive inference ➔ two level model ➔ results
Summary

The validity of inference rule applications is context sensitive.
Summary

The validity of inference rule applications is context sensitive

Our model outperforms both word-level and topic-level baselines on an unbiased test set
Summary

The validity of inference rule applications is context sensitive.

Our model outperforms both word-level and topic-level baselines on an unbiased test set.

See our rule resource at:
http://u.cs.biu.ac.il/~nlp/downloads/wt-rules.html

Future work - lexical inference.
Thank you!

The validity of inference rule applications is context sensitive.

Our model outperforms both word-level and topic-level baselines on an unbiased test set.

See our rule resource at:
http://u.cs.biu.ac.il/~nlp/downloads/wt-rules.html

Future work - lexical inference

topic level context sensitivity
word level similarity