Path-based vs. Distributional Information in Recognizing Lexical Semantic Relations

Vered Shwartz and Ido Dagan

Bar-Ilan University

December 12, 2016
Lexical Semantic Relations

- Some interesting lexical semantic relations:
  - Synonymy: \((\text{elevator}, \text{lift})\)
  - Hypernymy (Hyponymy): \((\text{green}, \text{color}), (\text{Obama}, \text{president})\)
  - Meronymy (Holonomy): \((\text{London}, \text{England}), (\text{hand}, \text{body})\)
  - Antonymy: \((\text{cold}, \text{hot})\)
  - etc.
Some interesting lexical semantic relations:

- **Synonymy**: (elevator, lift)
- **Hypernymy (Hyponymy)**: (green, color), (Obama, president)
- **Meronymy (Holonymy)**: (London, England), (hand, body)
- **Antonymy**: (cold, hot)
- etc.

Relations are used to infer one word from another in certain contexts:

- I ate an apple → I ate a fruit
- I hate fruit → I hate apples
- I visited Tokyo → I visited Japan
- I left Tokyo / I left Japan
Recognizing Lexical Semantic Relations

- Given two terms, \( x \) and \( y \), decide what is the semantic relation that holds between them (if any)
  - in some senses of \( x \) and \( y \)
  - e.g. both \textit{fruit} and \textit{company} are hypernyms of \textit{apple}
Example Motivation - Recognizing Textual Entailment

Text
A boy is hitting a baseball

Hypotheses

1. A child is hitting a baseball
   \[\text{ENTAILMENT} \quad \text{hypernym: (boy, child)}\]

2. A boy is missing a baseball
   \[\text{CONTRADICTION} \quad \text{antonym: (hitting, missing)}\]

3. A girl is hitting a baseball
   \[\text{NEUTRAL} \quad \text{co-hyponym: (boy, girl)}\]

Vered Shwartz (Bar-Ilan University)
Example Motivation - Recognizing Textual Entailment

Text

A boy is hitting a baseball

Hypotheses

1. A **child** is hitting a baseball ⇒ **ENTAILMENT**: hypernym: (boy, child)
Example Motivation - Recognizing Textual Entailment

Text
A boy is hitting a baseball

Hypotheses
1. A child is hitting a baseball ⇒ **ENTAILMENT**: hypernym: (boy, child)
2. A boy is missing a baseball ⇒ **CONTRADICTION**: antonym: (hitting, missing)
Example Motivation - Recognizing Textual Entailment

Text
A boy is hitting a baseball

Hypotheses
1. A **child** is hitting a baseball ⇒ **ENTAILMENT**: hypernym: (boy, child)
2. A boy is **missing** a baseball ⇒ **CONTRADICTION**: antonym: (hitting, missing)
3. A **girl** is hitting a baseball ⇒ **NEUTRAL**: co-hyponym: (boy, girl)
What’s in the talk?

- Overview of prior path-based and distributional methods for semantic relation classification

- Analysis of the contribution of each information source to the classification

Vered Shwartz (Bar-Ilan University)
What’s in the talk?

- Overview of **prior** path-based and distributional **methods** for semantic relation classification
- ...LSTM-based **architecture** for semantic relation classification (sorry!)
What’s in the talk?

- Overview of **prior** path-based and distributional **methods** for semantic relation classification
- ...LSTM-based **architecture** for semantic relation classification (sorry!)
- **Analysis** of the contribution of each information source to the classification
Prior Methods
Semantic Relation Classification

Corpus-based Methods

- Distributional
- Path-based
Distributional Methods

- Corpus-based Methods
  - Distributional
  - Path-based
Recognize the relation between $x$ and $y$ based on their *separate* occurrences in the corpus
- Namely: based on their neighbor distributions [Harris, 1954]
Distributional Methods

- Recognize the relation between $x$ and $y$ based on their *separate* occurrences in the corpus
  - Namely: based on their neighbor distributions [Harris, 1954]

- In earlier methods, words are represented as count-based vectors:

<table>
<thead>
<tr>
<th></th>
<th>elevator</th>
<th>lift</th>
<th>up</th>
<th>stairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0.56</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0.55</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0.89</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0.91</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>...</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>...</td>
<td></td>
<td>0</td>
</tr>
</tbody>
</table>
Distributional Methods

- Recognize the relation between $x$ and $y$ based on their *separate* occurrences in the corpus
  - Namely: based on their neighbor distributions [Harris, 1954]

- In earlier methods, words are represented as count-based vectors:

<table>
<thead>
<tr>
<th></th>
<th>elevator</th>
<th>lift</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td></td>
<td>0.89</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>0.91</td>
</tr>
<tr>
<td></td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td>0</td>
</tr>
</tbody>
</table>

  - Recent years: words are represented using low-dimensional word embeddings [Mikolov et al., 2013, Pennington et al., 2014]
(x, y) term-pairs are represented as a feature vector, based of the terms’ embeddings:
- Concatenation $\vec{x} \oplus \vec{y}$ [Baroni et al., 2012]
- Difference $\vec{y} - \vec{x}$ [Roller et al., 2014, Weeds et al., 2014]
$(x, y)$ term-pairs are represented as a feature vector, based on the terms’ embeddings:

- Concatenation $\vec{x} \oplus \vec{y}$ [Baroni et al., 2012]
- Difference $\vec{y} - \vec{x}$ [Roller et al., 2014, Weeds et al., 2014]

A classifier is trained to predict the semantic relation between $x$ and $y$
Supervised Distributional Methods

- \((x, y)\) term-pairs are represented as a feature vector, based of the terms’ embeddings:
  - Concatenation \(\vec{x} \oplus \vec{y}\) [Baroni et al., 2012]
  - Difference \(\vec{y} - \vec{x}\) [Roller et al., 2014, Weeds et al., 2014]
- A classifier is trained to predict the semantic relation between \(x\) and \(y\)
- Achieved very good results on common datasets
[Levy et al., 2015]: supervised distributional method do not learn the relation between $x$ and $y$
[Levy et al., 2015]: supervised distributional method do not learn the relation between $x$ and $y$

- Instead, they memorize single words that occur in the same relation
- e.g. that *fruit* or *animal* are prototypical hypernyms
Supervised Distributional Methods

- [Levy et al., 2015]: supervised distributional method do not learn the relation between $x$ and $y$
  - Instead, they memorize single words that occur in the same relation
  - e.g. that *fruit* or *animal* are prototypical hypernyms
- Recent work suggests these methods actually do more than memorize [Roller and Erk, 2016]
Supervised Distributional Methods

- [Levy et al., 2015]: supervised distributional method do not learn the relation between $x$ and $y$
  - Instead, they memorize single words that occur in the same relation
  - e.g. that *fruit* or *animal* are prototypical hypernyms
- Recent work suggests these methods actually do more than memorize [Roller and Erk, 2016]
- But they still provide only the *prior* of $x$ or $y$ to fit the relation [Shwartz et al., 2016]
Path-based Methods

Corpus-based Methods

Distributional

Path-based
Path-based Methods

- Recognize the relation between $x$ and $y$ based on their joint occurrences in the corpus

[Hearst, 1992] defined a set of patterns that indicate hypernymy, e.g.:

- $X$ or other $Y$
- $X$ is a $Y$
- $Y$, including $X$
Path-based Methods

- Recognize the relation between \( x \) and \( y \) based on their *joint* occurrences in the corpus
- Patterns connecting \( x \) and \( y \) may indicate the semantic relation that holds between them

[Hearst, 1992] defined a set of patterns that indicate hypernymy, e.g.:
- \( X \) or other \( Y \)
- \( X \) is a \( Y \)
- \( Y \), including \( X \)

Vered Shwartz (Bar-Ilan University)

LexNET

CogALEx 2016
Path-based Methods

- Recognize the relation between $x$ and $y$ based on their *joint* occurrences in the corpus.
- Patterns connecting $x$ and $y$ may indicate the semantic relation that holds between them.
- [Hearst, 1992] defined a set of patterns that indicate hypernymy, e.g.:
  - $X$ or other $Y$
  - $X$ is a $Y$
  - $Y$, including $X$
[Snow et al., 2004] represented patterns as the dependency paths connecting the words:

\[
\text{apple} \quad \text{is} \quad \text{a} \quad \text{fruit}
\]

Methods inspired by [Snow et al., 2004] were broadly adopted for semantic relation classification [Snow et al., 2006, Turney, 2006, Riedel et al., 2013].
[Snow et al., 2004] represented patterns as the dependency paths connecting the words:

They learned a hypernymy classifier using paths as features.
Path-based Methods

- [Snow et al., 2004] represented patterns as the dependency paths connecting the words:

  apple (NOUN) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ is (VERB) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ a (DET) fruit (NOUN)

- They learned a hypernymy classifier using paths as features
- Methods inspired by [Snow et al., 2004] were broadly adopted for semantic relation classification
  [Snow et al., 2006, Turney, 2006, Riedel et al., 2013]
However, this method suffers from relatively low recall:
Path-based Methods

However, this method suffers from relatively low recall:
1. $x$ and $y$ must occur together in the corpus (enough times!)
However, this method suffers from relatively low recall:

1. $x$ and $y$ must occur together in the corpus (enough times!)
2. paths connecting $x$ and $y$ must occur enough times in the corpus
However, this method suffers from relatively low recall:

1. $x$ and $y$ must occur together in the corpus (enough times!)
2. paths connecting $x$ and $y$ must occur enough times in the corpus

[Necșulescu et al., 2015] tried to overcome (1) by concatenating paths – e.g. by a connecting word $z$. 
Path-based Methods

- However, this method suffers from relatively low recall:
  1. $x$ and $y$ must occur together in the corpus (enough times!)
  2. paths connecting $x$ and $y$ must occur enough times in the corpus
- [Necșulescu et al., 2015] tried to overcome (1) by concatenating paths – e.g. by a connecting word $z$.
  - Performance improved, but the method was still outperformed by a distributional baseline
Path-based Methods

- However, this method suffers from relatively low recall:
  1. $x$ and $y$ must occur together in the corpus (enough times!)
  2. paths connecting $x$ and $y$ must occur enough times in the corpus

- [Necșulescu et al., 2015] tried to overcome (1) by concatenating paths – e.g. by a connecting word $z$.
  - Performance improved, but the method was still outperformed by a distributional baseline

- [PATTY] [Nakashole et al., 2012] tried to overcome (2) by generalized paths, e.g.: $X$ is defined/described as $Y \Rightarrow X$ is VERB as $Y$
Path-based Methods

- However, this method suffers from relatively low recall:
  1. $x$ and $y$ must occur together in the corpus (enough times!)
  2. paths connecting $x$ and $y$ must occur enough times in the corpus

- [Necșulescu et al., 2015] tried to overcome (1) by concatenating paths – e.g. by a connecting word $z$.
  - Performance improved, but the method was still outperformed by a distributional baseline

- PATTY [Nakashole et al., 2012] tried to overcome (2) by generalized paths, e.g.: $X$ is defined/described as $Y \Rightarrow X$ is VERB as $Y$
  - Syntactic generalizations make semantic mistakes: $X$ is rejected as $Y$
Integrated Methods

Corpus-based Methods

Distributional

Path-based

Integrated Methods
Path-based and distributional information sources are considered complementary in recognizing semantic relations.
Path-based and distributional information sources are considered complementary in recognizing semantic relations.

Classifiers with path-based and distributional features:
[Mirkin et al., 2006, Pavlick et al., 2015]
Path-based and distributional information sources are considered complementary in recognizing semantic relations.

Classifiers with path-based and distributional features:
- [Mirkin et al., 2006, Pavlick et al., 2015]
- HypeNET [Shwartz et al., 2016]: integrated path-based and distributional method for hypernymy detection
Path-based and distributional information sources are considered complementary in recognizing semantic relations.

Classifiers with path-based and distributional features:
[Mirkin et al., 2006, Pavlick et al., 2015]

HypeNET [Shwartz et al., 2016]: integrated path-based and distributional method for hypernymy detection
  - Improved neural path representation.
Path-based and distributional information sources are considered complementary in recognizing semantic relations.

Classifiers with path-based and distributional features:
[Mirkin et al., 2006, Pavlick et al., 2015]

HypeNET [Shwartz et al., 2016]: integrated path-based and distributional method for hypernymy detection
- Improved neural path representation
- Integrated with distributional features, and trained jointly
Path-based and distributional information sources are considered complementary in recognizing semantic relations.

Classifiers with path-based and distributional features: [Mirkin et al., 2006, Pavlick et al., 2015]

HypeNET [Shwartz et al., 2016]: integrated path-based and distributional method for hypernymy detection
  - Improved neural path representation
  - Integrated with distributional features, and trained jointly
  - ...works well also for multiple semantic relations (LexNET)
LexNET Architecture
An edge is a concatenation of 4 component vectors:

- dependent lemma
- dependent POS
- dependency label
- direction

Edges are fed sequentially to an LSTM to get the path embedding:

\[ \vec{v} \]
An edge is a concatenation of 4 component vectors:

- dependent lemma
- dependent POS
- dependency label
- direction

Edges are fed sequentially to an LSTM to get the path embedding:

\[
\vec{v}_{paths(x,y)}
\]

Embeddings:
- lemma
- POS
- dependency label
- direction

average pooling
Classification Models

1. Path-based
2. Distributional
3. Integrated
A classifier is trained on the path embedding $\vec{v}_{paths(x,y)}$:

$$V_{xy}$$

$$(x, y)$$ classification (softmax)
A classifier is trained on the concatenation of $x$ and $y$’s word embeddings $[\vec{v}_w^x, \vec{v}_w^y]$:

$$\vec{v}_{xy}$$

Vered Shwartz (Bar-Ilan University)
A classifier is trained on the concatenation of the path embedding $\vec{v}_{paths}(x,y)$, and $x$ and $y$'s word embeddings $[\vec{v}_w_x, \vec{v}_w_y]$. 
Integrated Model (LexNET)

- A classifier is trained on the concatenation of the path embedding $\vec{v}_{paths(x,y)}$, and $x$ and $y$’s word embeddings $[\vec{v}_{wx}, \vec{v}_{wy}]$.
- (HypeNET for multiple relations):
Non-linear Distributional Model ($DS_h$)

- [Levy et al., 2015]: linear classifiers incapable of learning interactions between $x$ and $y$’s features
Non-linear Distributional Model ($DS_h$)

- [Levy et al., 2015]: linear classifiers incapable of learning interactions between $x$ and $y$’s features
- How about adding non-linear expressive power? (i.e. hidden layer...)
  \[ v_{xy} \]
Similarly, we add a hidden layer to the integrated network:
Analysis
We tested our models on common semantic relations datasets:

- **K&H+N** [Necșulescu et al., 2015]
- **BLESS** [Baroni and Lenci, 2011]
- **ROOT09** [Santus et al., 2016]
- **EVALution** [Santus et al., 2015]
We tested our models on common semantic relations datasets:

- **K&H+N** [Necșulescu et al., 2015]
- **BLESS** [Baroni and Lenci, 2011]
- **ROOT09** [Santus et al., 2016]
- **EVALution** [Santus et al., 2015]

Each dataset contains several semantic relations, among:

- hypernymy
- meronymy
- co-hyponymy
- event
- attribute
- synonymy
- antonymy
- random
LexNET outperforms individual path-based and distributional methods

<table>
<thead>
<tr>
<th>Method</th>
<th>K&amp;H+N</th>
<th>BLESS</th>
<th>ROOT09</th>
<th>EVALution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>F₁</td>
<td>P</td>
</tr>
<tr>
<td>Path-based (PB)</td>
<td>0.713</td>
<td>0.604</td>
<td>0.55</td>
<td>0.759</td>
</tr>
<tr>
<td>Distributional (DSₖ)</td>
<td>0.983</td>
<td>0.984</td>
<td>0.983</td>
<td>0.891</td>
</tr>
<tr>
<td>Integrated (LexNET)</td>
<td>0.985</td>
<td>0.986</td>
<td>0.985</td>
<td>0.894</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.601</td>
</tr>
</tbody>
</table>
LexNET outperforms individual path-based and distributional methods

<table>
<thead>
<tr>
<th>method</th>
<th>K&amp;H+N</th>
<th>BLESS</th>
<th>ROOT09</th>
<th>EVALution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>F₁</td>
<td>P</td>
</tr>
<tr>
<td>Path-based (PB)</td>
<td>0.713</td>
<td>0.604</td>
<td>0.55</td>
<td>0.759</td>
</tr>
<tr>
<td>Distributional (DS₉₅)</td>
<td>0.983</td>
<td>0.984</td>
<td>0.983</td>
<td>0.891</td>
</tr>
<tr>
<td>Integrated (LexNET)</td>
<td>0.985</td>
<td>0.986</td>
<td>0.985</td>
<td><strong>0.894</strong></td>
</tr>
</tbody>
</table>

Vered Shwartz (Bar-Ilan University)
LexNET outperforms individual path-based and distributional methods

<table>
<thead>
<tr>
<th>method</th>
<th>K&amp;H+N</th>
<th></th>
<th>BLESS</th>
<th></th>
<th>ROOT09</th>
<th></th>
<th>EVALution</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>F₁</td>
<td>P</td>
<td>R</td>
<td>F₁</td>
<td>P</td>
<td>R</td>
</tr>
<tr>
<td>Path-based (PB)</td>
<td>0.713</td>
<td>0.604</td>
<td>0.55</td>
<td>0.759</td>
<td>0.756</td>
<td>0.755</td>
<td>0.788</td>
<td>0.789</td>
</tr>
<tr>
<td>Distributional (DS₉)</td>
<td>0.983</td>
<td>0.984</td>
<td>0.983</td>
<td>0.891</td>
<td>0.889</td>
<td>0.889</td>
<td>0.712</td>
<td>0.721</td>
</tr>
<tr>
<td>Integrated (LexNET)</td>
<td>0.985</td>
<td>0.986</td>
<td>0.985</td>
<td>0.894</td>
<td>0.893</td>
<td>0.893</td>
<td>0.813</td>
<td>0.814</td>
</tr>
</tbody>
</table>
LexNET outperforms individual path-based and distributional methods

<table>
<thead>
<tr>
<th>method</th>
<th>K&amp;H+N</th>
<th>BLESS</th>
<th>ROOT09</th>
<th>EVALution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Path-based (PB)</td>
<td>P</td>
<td>R</td>
<td>F₁</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>0.713</td>
<td>0.604</td>
<td>0.55</td>
<td>0.759</td>
</tr>
<tr>
<td>Integrated (LexNET)</td>
<td>P</td>
<td>R</td>
<td>F₁</td>
<td>P</td>
</tr>
<tr>
<td></td>
<td>0.985</td>
<td>0.986</td>
<td>0.985</td>
<td>0.894</td>
</tr>
<tr>
<td>Distributional (DSₜ)</td>
<td>P</td>
<td>R</td>
<td>F₁</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.983</td>
<td>0.984</td>
<td>0.983</td>
<td>0.891</td>
</tr>
</tbody>
</table>

0.985 0.986 0.985 0.894 0.893 0.893 0.813 0.814 0.813 0.57 0.573 0.571 0.601 0.607 0.6
Path-based contribution over distributional info is small in some datasets:

<table>
<thead>
<tr>
<th>method</th>
<th>K&amp;H+N</th>
<th></th>
<th>BLESS</th>
<th></th>
<th>ROOT09</th>
<th></th>
<th>EVALution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>F₁</td>
<td>P</td>
<td>R</td>
<td>F₁</td>
<td>P</td>
</tr>
<tr>
<td>Path-based (PB)</td>
<td>0.713</td>
<td>0.604</td>
<td>0.55</td>
<td>0.759</td>
<td>0.756</td>
<td>0.755</td>
<td>0.788</td>
</tr>
<tr>
<td>Distributional (DSₙ)</td>
<td>0.983</td>
<td>0.984</td>
<td>0.983</td>
<td>0.891</td>
<td>0.889</td>
<td>0.889</td>
<td>0.712</td>
</tr>
<tr>
<td>Integrated (LexNET)</td>
<td>0.985</td>
<td>0.986</td>
<td>0.985</td>
<td>0.894</td>
<td>0.893</td>
<td>0.893</td>
<td>0.813</td>
</tr>
</tbody>
</table>
**Contribution is prominent in unbiased datasets (i.e. when lexical memorization is disabled)**

<table>
<thead>
<tr>
<th>Method</th>
<th>K&amp;H+N</th>
<th>BLESS</th>
<th>ROOT09</th>
<th>EVALution</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P</td>
<td>R</td>
<td>F&lt;sub&gt;1&lt;/sub&gt;</td>
<td>P</td>
</tr>
<tr>
<td>Path-based (PB)</td>
<td>0.713</td>
<td>0.604</td>
<td>0.55</td>
<td>0.759</td>
</tr>
<tr>
<td>Distributional (DS&lt;sub&gt;H&lt;/sub&gt;)</td>
<td>0.983</td>
<td>0.984</td>
<td>0.983</td>
<td>0.891</td>
</tr>
<tr>
<td>Integrated (LexNET)</td>
<td>0.985</td>
<td>0.986</td>
<td>0.985</td>
<td>0.894</td>
</tr>
</tbody>
</table>

**Path-based models don’t memorize words! e.g. classify** *(cat, fruit)* **and** *(apple, animal)* **as randoms**
Path-based contribution over distributional info is prominent in the following scenarios:

- \( x \) or \( y \) are polysemous, e.g. \textit{mero}:\textit{(piano, key)}.
Path-based contribution over distributional info is prominent in the following scenarios:

- \( x \) or \( y \) are polysemous, e.g. \textit{mero}:\textit{(piano, key)}.
- the relation is not prototypical, e.g. \textit{event}:\textit{(cherry, pick)}.
Path-based contribution over distributional info is prominent in the following scenarios:

- $x$ or $y$ are polysemous, e.g. $mero:(piano, key)$. 
- the relation is not prototypical, e.g. $event:(cherry, pick)$. 
- $x$ or $y$ are rare, e.g. $hyper:(mastodon, proboscidean)$.
Path-based contribution over distributional info is prominent in the following scenarios:

- $x$ or $y$ are polysemous, e.g. `mero:(piano, key)`.
- The relation is not prototypical, e.g. `event:(cherry, pick)`.
- $x$ or $y$ are rare, e.g. `hyper:(mastodon, proboscidean)`.

Thanks to the path representation, such relations are captured even with a single meaningful co-occurrence of $x$ and $y$.
All models are bad in recognizing synonyms and antonyms.
All models are bad in recognizing synonyms and antonyms.

Path-based:
- Synonyms do not tend to occur together
All models are bad in recognizing synonyms and antonyms.

**Path-based:**
- Synonyms do not tend to occur together
- Antonyms occur in similar paths as co-hyponyms: *hot and cold, cats and dogs*
All models are bad in recognizing synonyms and antonyms.

- **Path-based:**
  - Synonyms do not tend to occur together
  - Antonyms occur in similar paths as co-hyponyms: *hot and cold, cats and dogs*

- **Distributional:**
  - Synonyms and antonyms occur in similar contexts:
    - “go down in the elevator/lift”, “it is hot/cold today”
Analysis of Semantic Relations

- All models are bad in recognizing synonyms and antonyms.
  - **Path-based:**
    - Synonyms do not tend to occur together
    - Antonyms occur in similar paths as co-hyponyms: *hot* and *cold*, *cats* and *dogs*
  - **Distributional:**
    - Synonyms and antonyms occur in similar contexts:
      - “go down in the elevator/lift”, “it is hot/cold today”
    - Problem not yet solved! Can we integrate additional information sources? :)

Vered Shwartz (Bar-Ilan University)
We presented LexNET, an adaptation of HypeNET [Shwartz et al., 2016] for multiple semantic relations.
We presented LexNET, an adaptation of HypeNET [Shwartz et al., 2016] for multiple semantic relations. LexNET outperforms the individual path-based / distributional models.
Recap

- We presented LexNET, an adaptation of HypeNET [Shwartz et al., 2016] for multiple semantic relations.
- LexNET outperforms the individual path-based / distributional models.
- The distributional source is dominant across most datasets.
Recap

- We presented LexNET, an adaptation of HypeNET [Shwartz et al., 2016] for multiple semantic relations.
- LexNET outperforms the individual path-based / distributional models.
- The distributional source is dominant across most datasets.
  - However, it learns the characteristics of a **single** word.
We presented LexNET, an adaptation of HypeNET [Shwartz et al., 2016] for multiple semantic relations.

LexNET outperforms the individual path-based / distributional models.

The distributional source is dominant across most datasets.

However, it learns the characteristics of a single word.

Path-based information always contributes to the classification.
We presented LexNET, an adaptation of HypeNET [Shwartz et al., 2016] for multiple semantic relations. LexNET outperforms the individual path-based / distributional models. The distributional source is dominant across most datasets. However, it learns the characteristics of a single word. Path-based information always contributes to the classification. It captures the relation between the words.
Recap

- We presented LexNET, an adaptation of HypeNET [Shwartz et al., 2016] for multiple semantic relations.
- LexNET outperforms the individual path-based / distributional models.
- The distributional source is dominant across most datasets.
  - However, it learns the characteristics of a single word.
- Path-based information always contributes to the classification.
  - It captures the relation between the words.

Thank you!


