Distributional Semantics

Advanced Machine Learning for NLP
Jordan Boyd-Graber

SLIDES ADAPTED FROM YOAV GOLDBERG AND OMER LEVY
Beyond word2vec

- word2vec is factorizing a word-context matrix.
- The content of this matrix affects the resulting similarities.
- word2vec allows you to specify a *window size*.
- But what about other types of contexts?

- Example: *dependency contexts* (Levy and Dagan, ACL 2014)
Bag of Words (BoW) Context

Australian scientist discovers star with telescope
Bag of Words (BoW) Context

Australian scientist discovers star with telescope
Australian scientist discovers star with telescope

Australian scientist **discovers** star with telescope
Australian scientist discovers star with telescope
Australian scientist discovers star with telescope

Syntactic Dependency Context

nsubj > discovers > dobj

prep_with
## Embedding Similarity with Different Contexts

<table>
<thead>
<tr>
<th>Target Word</th>
<th>Bag of Words (k=5)</th>
<th>Dependencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hogwarts</td>
<td>Dumbledore, hallows, half-blood, Malfoy, Snape</td>
<td>Sunnydale, Collinwood, Calarts, Greendale, Millfield</td>
</tr>
<tr>
<td>(Harry Potter’s school)</td>
<td></td>
<td>Related to Harry Potter, Schools</td>
</tr>
</tbody>
</table>
Embedding Similarity with Different Contexts

<table>
<thead>
<tr>
<th>Target Word</th>
<th>Bag of Words (k=5)</th>
<th>Dependencies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Turing (computer scientist)</td>
<td>nondeterministic</td>
<td>Pauling</td>
</tr>
<tr>
<td></td>
<td>non-deterministic</td>
<td>Hotelling</td>
</tr>
<tr>
<td></td>
<td>computability</td>
<td>Heting</td>
</tr>
<tr>
<td></td>
<td>deterministic</td>
<td>Lessing</td>
</tr>
<tr>
<td></td>
<td>finite-state</td>
<td>Hamming</td>
</tr>
<tr>
<td></td>
<td><strong>Related to computability</strong></td>
<td><strong>Scientists</strong></td>
</tr>
</tbody>
</table>

- Pauling
- Hotelling
- Heting
- Lessing
- Hamming

**Scientists**
### Embedding Similarity with Different Contexts

<table>
<thead>
<tr>
<th>Target Word</th>
<th>Bag of Words (k=5)</th>
<th>Dependencies</th>
</tr>
</thead>
</table>
| dancing (dance gerund) | singing  
dance  
dances  
dancers  
tap-dancing | singing  
rapping  
braedancing  
miming  
busking |

**Related to dance**

**Gerunds**

**Online Demo!**
Context matters

Choose the correct contexts for your application

• larger window sizes – more topical
• dependency relations – more functional
Context matters

Choose the correct contexts for your application

- larger window sizes – more topical
- dependency relations – more functional
- only noun-adjective relations
- only verb-subject relations
Context matters

Choose the correct contexts for your application

- larger window sizes – more topical
- dependency relations – more functional
- only noun-adjective relations
- only verb-subject relations
- context: time of the current message
- context: user who wrote the message
Context matters

Choose the correct contexts for your application

- larger window sizes – more topical
- dependency relations – more functional
- only noun-adjective relations
- only verb-subject relations
- context: time of the current message
- context: user who wrote the message
- ...
- the sky is the limit
## Distributional Semantics

- Words in similar contexts have similar meanings.
- Represent a word by the contexts it appears in.
- But what is a context?

## Neural Models (word2vec)

- Represent each word as dense, low-dimensional vector.
- Same intuitions as in distributional vector-space models.
- Efficient to run, scales well, modest memory requirement.
- Dense vectors are convenient to work with.
- Still helpful to think of the context types.