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)
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

Syntactic Dependency Context

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

Syntactic Dependency Context

- nsubj
- prep_with
- dobj

Australian scientist **discovers** star telescope
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 (Harry Potter’s school)</td>
<td>Dumbledore, hallows, half-blood, Malfoy, Snape</td>
<td>Sunnydale, Collinwood, Calarts, Greendale, Millfield</td>
</tr>
</tbody>
</table>

Related to Harry Potter Schools
## 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</td>
<td>nondeterministic</td>
<td>Pauling</td>
</tr>
<tr>
<td>(computer scientist)</td>
<td>non-deterministic computability</td>
<td>Hotelling</td>
</tr>
<tr>
<td></td>
<td>deterministic</td>
<td>Heting</td>
</tr>
<tr>
<td></td>
<td>finite-state</td>
<td>Lessing</td>
</tr>
</tbody>
</table>

Related to computability

Scientists

- Turing
- (computer scientist)
- nondeterministic
- non-deterministic
- computability
- deterministic
- finite-state

- Pauling
- Hotelling
- Heting
- Lessing
- Hamming

Scientists
Embedding Similarity with Different Contexts

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

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.