Still a Pain in the Neck
Evaluating Text Representations on Lexical Composition

Vered Shwartz and Ido Dagan

TACL 2019
Representing Phrases
How to represent a phrase $p = w_1 \ldots w_k$?

Most straightforward:

$$f(\vec{V}_{w_1}, \vec{V}_{w_2}, \ldots, \vec{V}_{w_k})$$
Representing Phrases
How to represent a phrase $p = w_1 \ldots w_k$?

Most straightforward:

$$f ( \vec{V}_{w_1}, \vec{V}_{w_2}, \ldots, \vec{V}_{w_k} )$$

“The whole is greater than the sum of its parts”
Representing Phrases
How to represent a phrase $p = w_1 \ldots w_k$?

Most straightforward:

$$f ( \vec{v}_{w_1}, \vec{v}_{w_2}, \ldots, \vec{v}_{w_k} )$$

“The whole is greater than the sum of its parts”

1. Meaning shift
2. Implicit meaning
Meaning Shift

A constituent word may be used in a non-literal way
Meaning Shift

A constituent word may be used in a non-literal way

VPC meanings differ from their verbs’ meanings
Implicit Meaning

Noun compounds

@_you_had_one_job1
Implicit Meaning

Noun compounds

Adjective-noun compositions

Please note, Cat milk does not come from cats

@_you_had_one_job1
Can existing representations address these phenomena?  
Probing Tasks

Simple tasks designed to test a single linguistic property  
[Adi et al., 2017, Conneau et al., 2018]
Probing Tasks
Representations

Standard / Contextualized

- word2vec
- GloVe
- fastText
- ELMo
- GPT
- BERT

Representation → Minimal Model → Prediction
Probing Tasks
Classifiers

1. Embed
2. Encode
3. Predict

Representation → Minimal Model → Prediction
Classifiers

1. **Embed**: each representation

2. **Encode**: none / biLSTM / self-attention

3. **Predict**:
   \[
   \vec{x} = \text{vector of target span, additional inputs}
   \]
   \[
   \vec{o} = \text{softmax}(W \cdot ReLU(\text{dropout}(h(\vec{x}))))
   \]
Probing Tasks

Tasks

Meaning shift / Implicit meaning

Representation → Minimal Model → Prediction

VPC Classification
LVC Classification
NC Literality
NC Relation
AN Relation
Phrase Type
(1) Meaning shift - human-like performance for contextualized
(2) Implicit meaning - far from humans
Meaning Shift Tasks
Verb-Particle Classification

Task Definition

VPC

We did get on together

Non-VPC

Which response did you get on that?
Results

We did get on together. Which response did you get on that?

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>VPC</td>
<td>23.6</td>
</tr>
<tr>
<td>Non-VPC</td>
<td></td>
</tr>
<tr>
<td>Majority</td>
<td>23.6</td>
</tr>
<tr>
<td>Human</td>
<td>93.8</td>
</tr>
<tr>
<td>word2vec</td>
<td>29.5</td>
</tr>
<tr>
<td>GloVe</td>
<td>60.5</td>
</tr>
<tr>
<td>fastText</td>
<td>45.5</td>
</tr>
<tr>
<td>ELMo</td>
<td>83.6</td>
</tr>
<tr>
<td>OpenAI GPT</td>
<td>81.4</td>
</tr>
<tr>
<td>BERT</td>
<td>90.0</td>
</tr>
</tbody>
</table>

Vered Shwartz  •  Evaluating Text Representations on Lexical Composition  •  November 7, 2019
We did get on together. Which response did you get on that?

Accuracy

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>VPC</td>
<td>23.6</td>
</tr>
<tr>
<td>word2vec</td>
<td>29.5</td>
</tr>
<tr>
<td>GloVe</td>
<td>60.5</td>
</tr>
<tr>
<td>fastText</td>
<td>45.5</td>
</tr>
<tr>
<td>ELMo</td>
<td>83.6</td>
</tr>
<tr>
<td>OpenAI GPT</td>
<td>81.4</td>
</tr>
<tr>
<td>BERT</td>
<td>90.0</td>
</tr>
<tr>
<td>Human</td>
<td>93.8</td>
</tr>
</tbody>
</table>
Results

We did get on together Which response did you get on that?

VPC
Non-VPC

Accuracy

<table>
<thead>
<tr>
<th>Method</th>
<th>Standard</th>
<th>Contextualized</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority</td>
<td>23.6</td>
<td></td>
</tr>
<tr>
<td>word2vec</td>
<td>29.5</td>
<td></td>
</tr>
<tr>
<td>GloVe</td>
<td>60.5</td>
<td></td>
</tr>
<tr>
<td>fastText</td>
<td>45.5</td>
<td></td>
</tr>
<tr>
<td>ELMo</td>
<td>83.6</td>
<td></td>
</tr>
<tr>
<td>OpenAI GPT</td>
<td>81.4</td>
<td></td>
</tr>
<tr>
<td>BERT</td>
<td>90.0</td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>93.8</td>
<td></td>
</tr>
</tbody>
</table>
Verb-Particle Classification Analysis

get on

make for

give in

take on
Noun Compound Literality

Task Definition

The crash course in litigation made me a better lawyer
Noun Compound Literality

Results

The crash course in litigation made me a better lawyer.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority</td>
<td>72.5</td>
</tr>
<tr>
<td>word2vec</td>
<td>72.5</td>
</tr>
<tr>
<td>GloVe</td>
<td>80.4</td>
</tr>
<tr>
<td>fastText</td>
<td>74.6</td>
</tr>
<tr>
<td>ELMo</td>
<td>79.7</td>
</tr>
<tr>
<td>OpenAI GPT</td>
<td>87.0</td>
</tr>
<tr>
<td>BERT</td>
<td>91.3</td>
</tr>
<tr>
<td>Human</td>
<td>91.0</td>
</tr>
</tbody>
</table>
Noun Compound Literality

Results

The crash course in litigation made me a better lawyer.

<table>
<thead>
<tr>
<th>Model</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority</td>
<td>72.5</td>
</tr>
<tr>
<td>word2vec</td>
<td>72.5</td>
</tr>
<tr>
<td>GloVe</td>
<td>80.4</td>
</tr>
<tr>
<td>fastText</td>
<td>74.6</td>
</tr>
<tr>
<td>Human</td>
<td>91.0</td>
</tr>
</tbody>
</table>
The crash course in litigation made me a better lawyer.
Noun Compound Literality
Detecting meaning shift → modeling meaning?

<table>
<thead>
<tr>
<th>ELMo</th>
<th>OpenAI GPT</th>
<th>BERT</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Queen and her husband were on a train <strong>trip</strong> from Sydney to Orange.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ride</td>
<td>to</td>
<td>travelling</td>
</tr>
<tr>
<td>carriage</td>
<td>headed</td>
<td>running</td>
</tr>
<tr>
<td>journey</td>
<td>heading</td>
<td>journey</td>
</tr>
<tr>
<td>heading</td>
<td>that</td>
<td>going</td>
</tr>
<tr>
<td>carrying</td>
<td>and</td>
<td>headed</td>
</tr>
</tbody>
</table>
## Noun Compound Literality

Detecting meaning shift → modeling meaning?

<table>
<thead>
<tr>
<th>ELMo</th>
<th>OpenAI GPT</th>
<th>BERT</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Queen and her husband were on a train <strong>trip</strong> from Sydney to Orange.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ride</td>
<td>to</td>
<td>travelling</td>
</tr>
<tr>
<td>carriage</td>
<td>headed</td>
<td>running</td>
</tr>
<tr>
<td>journey</td>
<td>heading</td>
<td>journey</td>
</tr>
<tr>
<td>heading</td>
<td>that</td>
<td>going</td>
</tr>
<tr>
<td>carrying</td>
<td>and</td>
<td>headed</td>
</tr>
</tbody>
</table>

Creating a guilt **trip** in another person may be considered to be psychological manipulation...

<table>
<thead>
<tr>
<th>tolerance</th>
<th>that</th>
<th>reaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>fest</td>
<td>so</td>
<td>feeling</td>
</tr>
<tr>
<td>avoidance</td>
<td><strong>trip</strong></td>
<td>attachment</td>
</tr>
<tr>
<td>onus</td>
<td>he</td>
<td>sensation</td>
</tr>
<tr>
<td>association</td>
<td>she</td>
<td>note</td>
</tr>
</tbody>
</table>
Noun Compound Literality
Non Decomposable Compounds

Substitutes for the entire phrase.

<table>
<thead>
<tr>
<th>ELMo</th>
<th>OpenAI GPT</th>
<th>BERT</th>
</tr>
</thead>
<tbody>
<tr>
<td>...I believe you are a snake oil salesman, a narcissist...</td>
<td></td>
<td></td>
</tr>
<tr>
<td>auto</td>
<td>in</td>
<td>oil</td>
</tr>
<tr>
<td>egg</td>
<td>and</td>
<td>pit</td>
</tr>
<tr>
<td>hunter</td>
<td>that</td>
<td>bite</td>
</tr>
<tr>
<td>rogue</td>
<td>charmer</td>
<td>jar</td>
</tr>
</tbody>
</table>
Implicit Meaning Tasks
Adjective-Noun Attributes

Task Definition

He receives warm support from his students.
Adjective-Noun Attributes

Results

He receives warm support from his students

Temperature

Emotionality

Accuracy

Majority

50.0

Human

86.4

50

100

Standard

Contextualized
He receives warm support from his students.

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority</td>
<td>50.0</td>
</tr>
<tr>
<td>word2vec</td>
<td>50.0</td>
</tr>
<tr>
<td>GloVe</td>
<td>50.0</td>
</tr>
<tr>
<td>fastText</td>
<td>53.8</td>
</tr>
<tr>
<td>OpenAI GPT</td>
<td>57.5</td>
</tr>
<tr>
<td>BERT</td>
<td>65.1</td>
</tr>
<tr>
<td>Human</td>
<td>86.4</td>
</tr>
</tbody>
</table>
Adjective-Noun Attributes

Results

He receives warm support from his students

Accuracy

Majority 50.0
word2vec 50.0
GloVe 50.0
fastText 53.8
ELMo 54.7
OpenAI GPT 57.5
BERT 65.1
Human 86.4

temperature
emotionality
Noun Compound Relations

Task Definition

The township is served by three access roads.

- Road forecasted for access season
- Road that makes access possible
Noun Compound Relations

Results

The township is served by three access roads.

Accuracy

0 50 100

Majority 50.0

Human 77.8

Standard

Contextualized

Road forecasted for access season

Road that makes access possible
The township is served by three access roads.

---

**Noun Compound Relations**

**Results**

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority</td>
<td>50.0</td>
</tr>
<tr>
<td>word2vec</td>
<td>50.0</td>
</tr>
<tr>
<td>GloVe</td>
<td>48.1</td>
</tr>
<tr>
<td>fastText</td>
<td>51.2</td>
</tr>
<tr>
<td>ELMo</td>
<td>54.3</td>
</tr>
<tr>
<td>OpenAI GPT</td>
<td>53.1</td>
</tr>
<tr>
<td>BERT</td>
<td>51.2</td>
</tr>
<tr>
<td>Human</td>
<td>77.8</td>
</tr>
</tbody>
</table>

Road forecasted for access season
Road that makes access possible
The township is served by three access roads.

Accuracy:
- Majority: 50.0%
- Word2vec: 50.0%
- GloVe: 48.1%
- fastText: 51.2%
- ELMo: 54.3%
- OpenAI GPT: 53.1%
- BERT: 51.2%
- Human: 77.8%

Contextualized vs. Standard text representations.
Future Directions
Can we learn phrase meanings like humans do?

- [Cooper, 1999]: how do L2 learners process idioms?
  - Infer from context: 28% (57% success rate)
  - Rely on literal meaning: 19% (22% success rate)
  - ...

Vered Shwartz · Evaluating Text Representations on Lexical Composition · November 7, 2019
Inferring from context

We need “extended” contexts
[Asl, 2013]: more successful idiom interpretation with extended contexts (stories)
Inferring from context

We need “extended” contexts
[Asl, 2013]: more successful idiom interpretation with extended contexts (stories)

We need richer context modeling
- Characters in the story
- Relationships between them
- Dialogues
- ...

Furious Meghan Markle says she won’t fall for dad’s ‘crocodile tears’ after he claimed ‘she’d be better off if he were dead’

FURIOUS Meghan Markle has said she won’t fall for her dad’s “crocodile tears” after he claimed “she’d be better off if he were dead”.

The Duchess of Sussex reportedly told pals Thomas Markle is using "emotional blackmail" to try and manipulate her but she’s had “enough already”.

Vered Shwartz • Evaluating Text Representations on Lexical Composition • November 7, 2019
Relying on literal meaning

“Robert knew he was robbing the cradle by dating a sixteen-year-old girl”

We need world knowledge

“Cradle is something you put the baby in”
Relying on literal meaning

“Robert knew he was robbing the cradle by dating a sixteen-year-old girl”

We need world knowledge

“Cradle is something you put the baby in”

We need to be able to reason

“You’re stealing a child from a mother”

“So robbing the cradle is like dating a really young person”

[Cooper, 1999]
Recap

- Detecting meaning shift is a *piece of cake*
Recap

Detecting meaning shift is a *piece of cake*
for contextualized word embeddings
Recap

- Detecting meaning shift is a *piece of cake*
  for contextualized word embeddings

- Modeling the shifted, rare sense is not a *walk in the park*
Recap

- Detecting meaning shift is a *piece of cake* for contextualized word embeddings

- Modeling the shifted, rare sense is not a *walk in the park*

- Modeling implicit information is a real *pain in the neck*
Recap

- Detecting meaning shift is a *piece of cake* for contextualized word embeddings.

- Modeling the shifted, rare sense is not a *walk in the park*.

- Modeling implicit information is a real *pain in the neck*.

Questions?
References


