The compositionality of neural networks: integrating symbolism and connectionism

Dieuwke Hupkes

Institute for Logic, Language and Computation
University of Amsterdam

May 6, 2019

The appropriateness of neural models

- "Modern approaches [. . .] do not explicitly formulate and execute compositional paths" (Johnson et al., 2017)

The appropriateness of neural models

- "Modern approaches [...] do not explicitly formulate and execute compositional paths" (Johnson et al., 2017)
- "Neural network models lack the abiltiy to extract systematic rules" (Lake and Baroni, 2018)

The appropriateness of neural models

- "Modern approaches [...] do not explicitly formulate and execute compositional paths" (Johnson et al., 2017)
- "Neural network models lack the abiltiy to extract systematic rules" (Lake and Baroni, 2018)
- "They do not learn in a compositional way" (Liška et al., 2018)

The appropriateness of neural models

- "Modern approaches [...] do not explicitly formulate and execute compositional paths" (Johnson et al., 2017)
- "Neural network models lack the abiltiy to extract systematic rules" (Lake and Baroni, 2018)
- "They do not learn in a compositional way" (Liška et al., 2018)
- "[...] neural networks are essentially very large correlation engines that hone in on any statisctical, potentially spurious pattern" (Hudson and Manning, 2018)

The appropriateness of neural models

- "Modern approaches [...] do not explicitly formulate and execute compositional paths" (Johnson et al., 2017)
- "Neural network models lack the abiltiy to extract systematic rules" (Lake and Baroni, 2018)
- "They do not learn in a compositional way" (Liška et al., 2018)
- "[...] neural networks are essentially very large correlation engines that hone in on any statisctical, potentially spurious pattern" (Hudson and Manning, 2018)
- Neural networks are data-hungry because they don't develop re-usable representations (almost everyone)

What is compositionality

The principle of compositionality

The meaning of a whole is a function of the meanings of the parts and of the way they are syntactically combined.

Partee (1995)

What is compositionality

What does it mean that neural networks are not compositional?

- They find different parts than we'd like them to
- They find different rules than we'd like them to
- They find other aspects of the data more salient
- They cannot represent hierarchy
- They favour memorising sequences over learning rules
- They are not getting the right signal from the data
- ...

The appropriateness of neural models

Our approach: "dissect" compositionality:

- Do models find the right parts and rules?

The appropriateness of neural models

Our approach: "dissect" compositionality:

- Do models find the right parts and rules?
- Do models use the parts and rules they finds **systematically**

The appropriateness of neural models

Our approach: “dissect” compositionality:

- Do models find the right parts and rules?
- Do models use the parts and rules they finds **systematically**
- Do models use the parts and rules they finds **productively**

The appropriateness of neural models

Our approach: “dissect” compositionality:

- Do models find the right parts and rules?
- Do models use the parts and rules they finds **systematically**
- Do models use the parts and rules they finds **productively**
- Do models compute **locally consistent** representations?

The appropriateness of neural models

Our approach: "dissect" compositionality:

- Do models find the right parts and rules?
- Do models use the parts and rules they finds **systematically**
- Do models use the parts and rules they finds **productively**
- Do models compute **locally consistent** representations?
- Do models allow **substitution** of synonyms?

The appropriateness of neural models

Our approach: "dissect" compositionality:

- Do models find the right parts and rules?
- Do models use the parts and rules they finds **systematically**
- Do models use the parts and rules they finds **productively**
- Do models compute **locally consistent** representations?
- Do models allow **substitution** of synonyms?
- Do models prefer **rules** or **exceptions**?

The rest of the team

Mathijs Mul

Verna Dankers

Elia Bruni

Data
PCFG SET

Unary functions: `reverse`, `swap`, `copy`, ...
Binary functions: `prepend`, `append`, `remove_first`, ...
Characters: `A`, `B`, `C`, ...

Data
PCFG SET

Unary functions: `reverse`, `swap`, `copy`, ...
Binary functions: `prepend`, `append`, `remove_first`, ...
Characters: `A`, `B`, `C`, ...

`reverse A B C`

Data
PCFG SET

Unary functions: `reverse`, `swap`, `copy`, ...
Binary functions: `prepend`, `append`, `remove_first`, ...
Characters: `A`, `B`, `C`, ...

`reverse A B C` $\Rightarrow$ `C B A`

Data
PCFG SET

Unary functions: `reverse`, `swap`, `copy`, ...
Binary functions: `prepend`, `append`, `remove_first`, ...
Characters: `A`, `B`, `C`, ...

```
reverse A B C          ⇒    C B A
append C B A , D E
```

Data

PCFG SET

Unary functions: `reverse`, `swap`, `copy`, ...
Binary functions: `prepend`, `append`, `remove_first`, ...
Characters: `A`, `B`, `C`, ...

```
reverse A B C        ⇒    C B A
append C B A , D E   ⇒    C B A D E
```

Data
PCFG SET

Unary functions: `reverse`, `swap`, `copy`, ...
Binary functions: `prepend`, `append`, `remove_first`, ...
Characters: `A`, `B`, `C`, ...

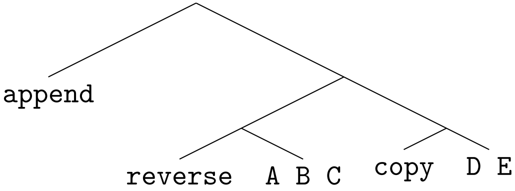
`reverse A B C` ⇒ `C B A`
`append C B A , D E` ⇒ `C B A D E`

`append reverse A B C , copy D E` ⇒ `C B A D E`

Data
PCFG SET

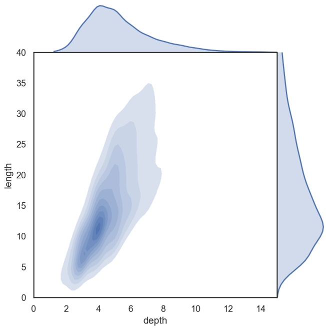
Unary functions: `reverse`, `swap`, `copy`, ...
Binary functions: `prepend`, `append`, `remove_first`, ...
Characters: `A`, `B`, `C`, ...

`append reverse A B C , copy D E` $\Rightarrow$ `C B A D E`

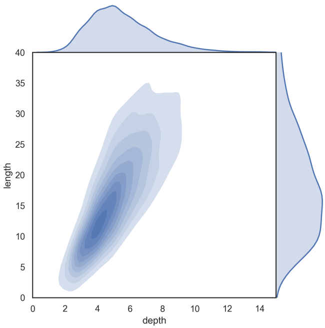


PCFG SET

Data Naturalisation



(a) PCFG SET



(b) WMT 2017

Figure: Distribution of sentence depth and length in the PCFG SET and WMT2017 data.

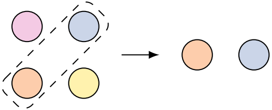
Models

1. **LSTMS2S** Recurrent encoder-decoder model with attention
2. **ConvS2S** Convolutional encoder and decoder with multistep attention
3. **Transformer** Fully attention based model

Results

Experiment	LSTMS2S	ConvS2S	Transformer
PCFG SET*	0.77 ± 0.01	0.84 ± 0.01	0.93 ± 0.01

Systematicity



Can models systematically recombine unseen pairs of functions?

Results

Systematicity

Experiment	LSTMS2S	ConvS2S	Transformer
PCFG SET*	0.77 ± 0.01	0.84 ± 0.01	0.93 ± 0.01
Systematicity*	0.51 ± 0.03	0.55 ± 0.01	0.70 ± 0.01

Localism

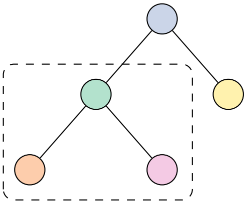


Figure: Localism

Do models build representations incrementally?

`append reverse A B C , copy D E`

$$\equiv$$

`append C B A , D E`

?

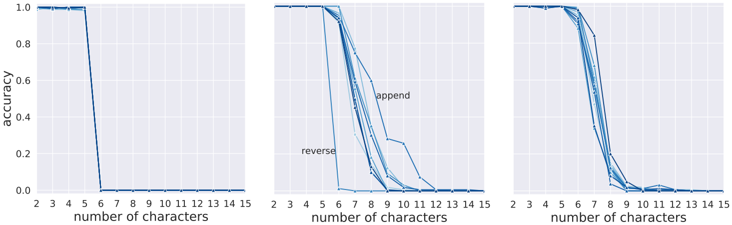
Results

Localism

Experiment	**LSTMS2S**	**ConvS2S**	**Transformer**
PCFG SET*	0.77 ± 0.01	0.84 ± 0.01	0.93 ± 0.01
Systematicity*	0.51 ± 0.03	0.55 ± 0.01	0.70 ± 0.01
Localism†	0.45 ± 0.01	0.57 ± 0.04	0.56 ± 0.03

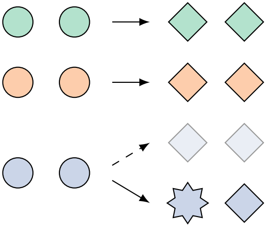
Results

Generality of representations



(a) LSTM2S (b) Conv2S (c) Transformer

Overgeneralisation



Do models overgeneralise during training?

Results

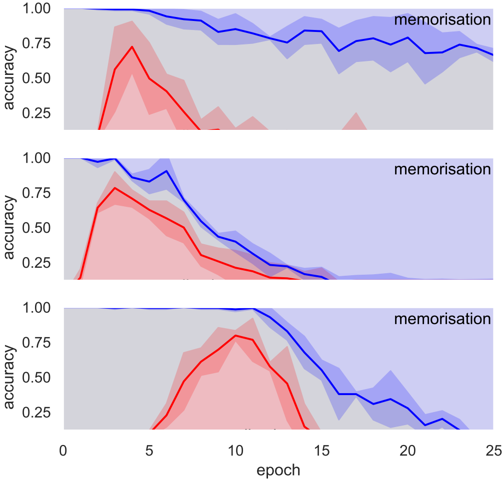
Experiment	LSTMS2S	ConvS2S	Transformer
PCFG SET*	0.77 ± 0.01	0.84 ± 0.01	0.93 ± 0.01
Systematicity*	0.51 ± 0.03	0.55 ± 0.01	0.70 ± 0.01
Localism†	0.45 ± 0.01	0.57 ± 0.04	0.56 ± 0.03
Overgeneralisation*	0.73 ± 0.18	0.78 ± 0.12	0.84 ± 0.02

Overgeneralisation profile

LSTM2S

ConvS2S

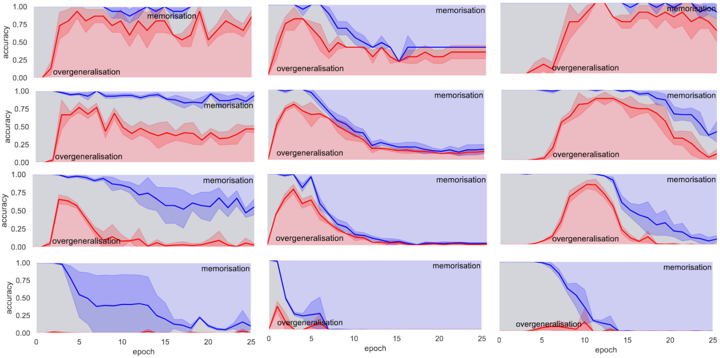
Transformer



Overgeneralisation

Different exception rates

Overgeneralisation profiles for exceptions occuring 0.01%, 0.05%, 0.1% and 0.5%



(a) LSTM2S (b) Conv2S (c) Transformer

The rest of the team

Mathijs Mul

Verna Dankers

Elia Bruni

References

Drew A. Hudson and Christopher D. Manning. Compositional attention networks for machine reasoning. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2018.

Justin Johnson, Bharath Hariharan, Laurens van der Maaten, Li Fei-Fei, C Lawrence Zitnick, and Ross Girshick. CLEVR: A diagnostic dataset for compositional language and elementary visual reasoning. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1988–1997. IEEE, 2017.

Brenden M. Lake and Marco Baroni. Still not systematic after all these years: On the compositional skills of sequence-to-sequence recurrent networks. In *ICLR 2018 workshop track*, 2018.

Adam Liška, Germán Kruszewski, and Marco Baroni. Memorize or generalize? searching for a compositional rnn in a haystack. In *ICML workshop Architectures and Evaluation for Generality, Autonomy and Progress in AI (AEGAP)*, 2018.

Barbara Partee. Lexical semantics and compositionality. *An invitation to cognitive science: Language*, 1:311–360, 1995.

Zoltán Gendler Szabó. Compositionality as supervenience. *Linguistics and Philosophy*, 23(5):475–505, 2000.