A Cognitively Plausible Adaptive Neural Language Model

Marten van Schijndel and Tal Linzen

May 12, 2018

Department of Cognitive Science, Johns Hopkins University

Subjects learn to expect vocabulary items and syntactic structures (Otten & Van Berkum, 2008; Fine et al., 2013)

Subjects learn to expect vocabulary items and syntactic structures (Otten & Van Berkum, 2008; Fine et al., 2013)

RRC: The soldiers warned about the dangers conducted the raid

Humans adapt to linguistic context

Subjects learn to expect vocabulary items and syntactic structures (Otten & Van Berkum, 2008; Fine et al., 2013)

RRC: The soldiers warned about the dangers conducted the raid

$$\text{P(RRC)} = \overset{\text{typical}}{0.008} \rightarrow \overset{\text{Fine et al.}}{0.50}$$

Subjects learn to expect vocabulary items and syntactic structures (Otten & Van Berkum, 2008; Fine et al., 2013)

RRC: The soldiers warned about the dangers conducted the raid

$\text{P(RRC)} = \overset{\text{typical}}{0.008} \rightarrow \overset{\text{Fine et al.}}{0.50}$

By end of experiment, subjects expected RRC more than at beginning

Adaptation studied in NLP

- Domain adaptation (Kuhn & de Mori, 1990; McClosky, 2010)

  News Model $\rightarrow$ Biomedical Text
- Handling unknown words (Grave et al., 2015)

  Learn new words from context
- Style adaptation (Jaech & Ostendorf, 2017)

  Lawyer A $\rightarrow$ Lawyer B

Adaptation studied in NLP

- Domain adaptation (Kuhn & de Mori, 1990; McClosky, 2010)
  News Model $\rightarrow$ Biomedical Text
- Handling unknown words (Grave et al., 2015)
  Learn new words from context
- Style adaptation (Jaech & Ostendorf, 2017)
  Lawyer A $\rightarrow$ Lawyer B

But can we model human adaptation?

Our proposed model

LSTM language model (Gives prob of next word in sequence)

Base Model: Trained on Wikipedia (90M words)
(Gulordava et al., 2018)

Our proposed model

LSTM language model (Gives prob of next word in sequence)

Base Model: Trained on Wikipedia (90M words)
(Gulordava et al., 2018)

Adaptation algorithm:

1. Test on a sentence
2. Update weights based on that sentence
3. Repeat on remaining sentences

Experiment 1:

Does adaptation improve prediction accuracy?

Accuracy Evaluation Measure: Perplexity

Perplexity:
How much probability mass is assigned to wrong words?

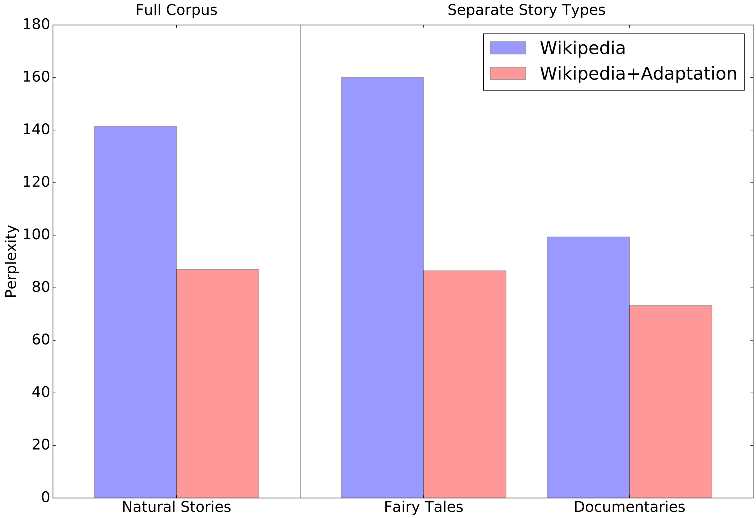
How surprised is the model by the data?
(Lower is better)

Accuracy Evaluation Data

Test data: Natural Stories Corpus (Futrell et al., 2017)

- 10 texts (485 sentences)
  - 7 Fairy Tales
  - 3 Documentaries

Accuracy Results



Experiment 2:
Are adaptive expectations human-like?

Psycholinguistic Evaluation Measure: Surprisal

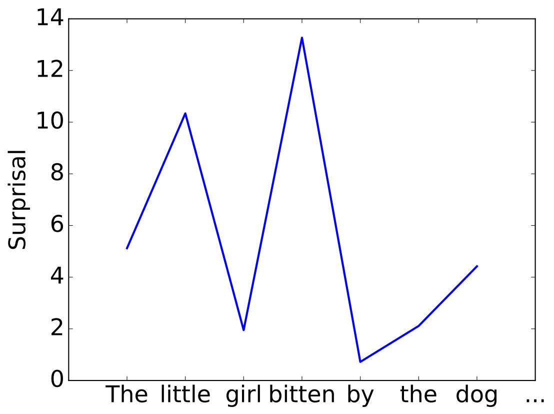
Reading times can be predicted with surprisal (Smith and Levy, 2013)

$$\text{Surprisal}(w_i) = -\log \text{P}(w_i \mid w_{1..i-1})$$

Psycholinguistic Evaluation Measure: Surprisal

Reading times can be predicted with surprisal (Smith and Levy, 2013)

$$\text{Surprisal}(w_i) = -\log \mathrm{P}(w_i \mid w_{1..i-1})$$



Psycholinguistic Evaluation Data: Reading Times

Test data: Natural Stories Corpus (Futrell et al., 2017)

Also contains self-paced reading times! ($N = 181$)

Test data: Natural Stories Corpus (Futrell et al., 2017)

Also contains self-paced reading times! ($N = 181$)

The ———————————————————

Psycholinguistic Evaluation Data: Reading Times

Test data: Natural Stories Corpus (Futrell et al., 2017)

Also contains self-paced reading times! ($N = 181$)

`--- boy ---------------------`

Test data: Natural Stories Corpus (Futrell et al., 2017)

Also contains self-paced reading times! ($N = 181$)

```
------- threew ---------------
```

Test data: Natural Stories Corpus (Futrell et al., 2017)

Also contains self-paced reading times! ($N = 181$)

`-------------- the ------------`

Psycholinguistic Evaluation Data: Reading Times

Test data: Natural Stories Corpus (Futrell et al., 2017)

Also contains self-paced reading times! ($N = 181$)

—————————— dog ———————

Test data: Natural Stories Corpus (Futrell et al., 2017)

Also contains self-paced reading times! ($N = 181$)

––––––––––––––––––––– a –––––

Test data: Natural Stories Corpus (Futrell et al., 2017)

Also contains self-paced reading times! ($N = 181$)

```
------------------------ ball.
```

Non-adaptive surprisal is a good predictor of reading times

	$\hat{\beta}$	$\hat{\sigma}$	t-value	
Sentence position	0.3592	0.5284	0.680	
Word length	6.3828	1.0034	6.361	***
Non-adaptive surprisal	8.4480	0.6294	13.422	***

Fixed effects of linear mixed regression

Adaptive surprisal is a better predictor of reading times

	$\hat{\beta}$	$\hat{\sigma}$	t-value	
Sentence position	0.2903	0.5310	0.547	
Word length	6.4266	1.0035	6.404	***
Non-adaptive surprisal	-0.8873	0.6754	-1.314	
Adaptive surprisal	8.7714	0.6764	12.968	***

Fixed effects of linear mixed regression

Experiment 3:
Does the model adapt to vocabulary, syntax, or both?

Generated 200 dative sentence pairs

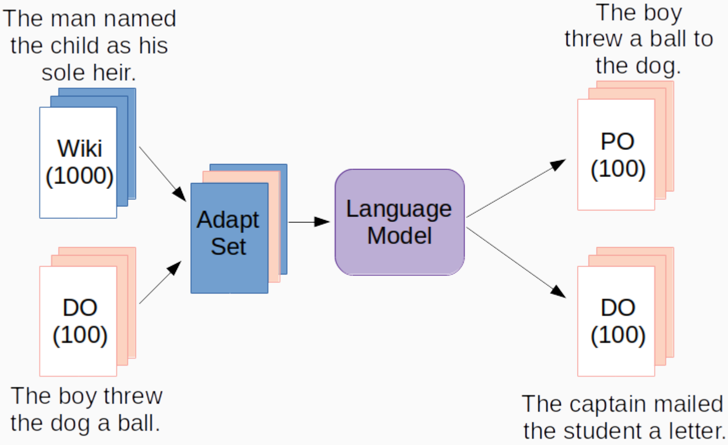
Prepositional Object (PO):

The boy threw the ball to the dog.

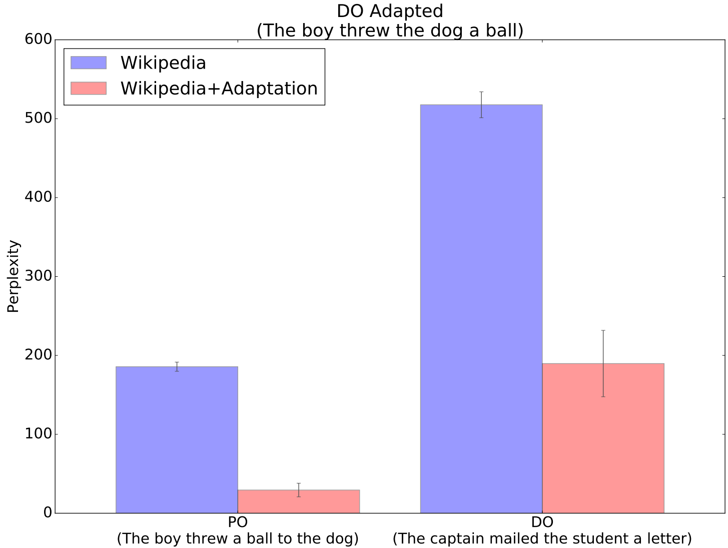
Double Object (DO):

The boy threw the dog the ball.

Dative evaluation paradigm



Model adapts to vocabulary *and* syntax



Our adaptive language model makes

- More accurate predictions
- More human-like predictions

than a non-adaptive language model.

- Adaptation driven by both vocabulary and syntax

Future directions:

- How sensitive are RT results to learning rate?
- Reproduce psycholinguistic adaptation results
- Compare adaptation mechanisms using human behavioral data

Thanks!

Model adapts to vocabulary *and* syntax