

What Gets Echoed?

Understanding the “Pointers” in Explanations of Persuasive Arguments

David Atkinson, Kumar Bhargav Srinivasan, and Chenhao Tan

{david.i.atkinson, kumar.srinivasan, chenhao.tan}@colorado.edu

Explanations are important.¹²³⁴⁵⁶⁷

¹(Keil, 2006)

²(Ribeiro et al., 2016)

³(Lipton, 2016)

⁴(Guidotti et al., 2019)

⁵(Miller, 2019)

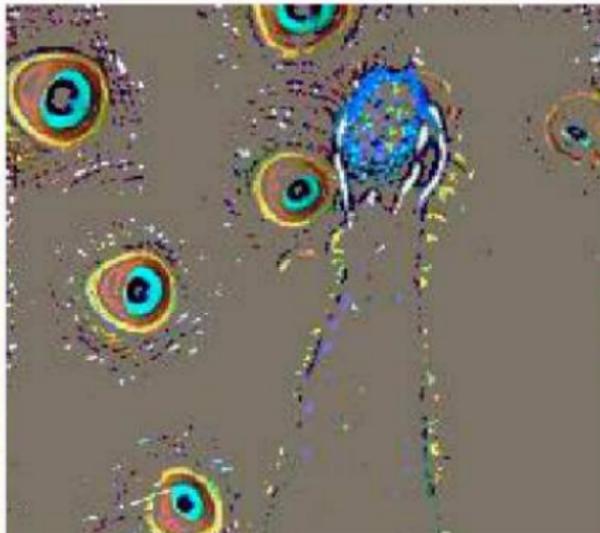
⁶(Doshi-Velez and Kim, 2019)

⁷...and so on.

What is this?



What is this?

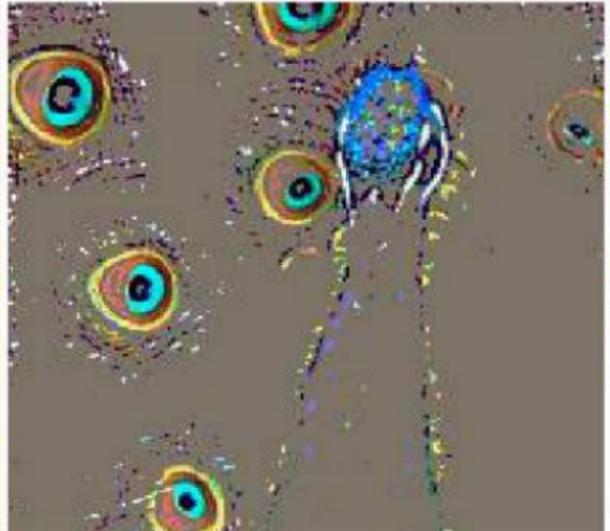


8

What is this?

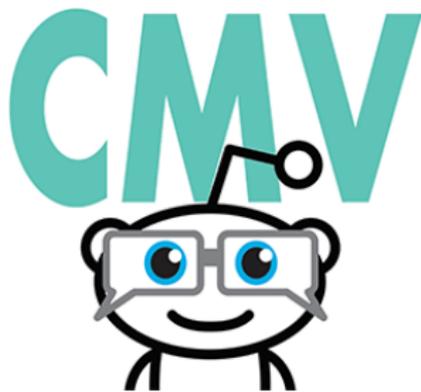


Explanandum



Explanation

What about **natural language** explanations?



What is [/r/changemyview](#)?

Virgina Heffernan, writing in *Wired*

“In a culture of brittle talking points that we guard with our lives, Change My View is a source of motion and surprise.”

r/ChangeMyView

OP

CMV: most hit music artists today are bad musicians

1 year ago by * (last edited 1 year ago) [ssbeluga](#)

They're super repetitive, the lyrics and painfully generic, and there's hardly ever anything new or challenging. And from what I understand most of these artists don't even write their own songs. It seems to me they're only popular because of who they are and how they look/perform. Can anyone convince me otherwise?

r/ChangeMyView

OP

CMV: most hit music artists today are bad musicians

1 year ago by * (last edited 1 year ago) [ssbeluga](#)

They're super repetitive, the lyrics and painfully generic, and there's hardly ever anything new or challenging. And from what I understand most of these artists don't even write their own songs. It seems to me they're only popular because of who they are and how they look/perform. Can anyone convince me otherwise?



[–] [kublahkoala](#) 229Δ 13 points 1 year ago

Music appreciation is a skill, and it's all about pattern recognition.

When we're children, we need songs that are really simple, with easy to recognize patterns. The younger we are, the simpler the songs. Toddlers like nursery rhymes. Teens like pop music.

You have to consider that music can be listened to for different purposes. Maybe you need something on in the background during a party, or while you study, or are trying to fall asleep, or work out. Pop music is really good in some of these situations, really bad in others.

r/ChangeMyView

OP

CMV: most hit music artists today are bad musicians

1 year ago by * (last edited 1 year ago) [ssbeluga](#)

They're super repetitive, the lyrics and painfully generic, and there's hardly ever anything new or challenging. And from what I understand most of these artists don't even write their own songs. It seems to me they're only popular because of who they are and how they look/perform. Can anyone convince me otherwise?



PC

[\[-\] kublahkoala](#) [229Δ](#) 13 points 1 year ago

Music appreciation is a skill, and it's all about pattern recognition.

When we're children, we need songs that are really simple, with easy to recognize patterns. The younger we are, the simpler the songs. Toddlers like nursery rhymes. Teens like pop music.

You have to consider that music can be listened to for different purposes. Maybe you need something on in the background during a party, or while you study, or are trying to fall asleep, or work out. Pop music is really good in some of these situations, really bad in others.

Explanation



[\[-\] ssbeluga](#) [\[S\]](#) 3 points 1 year ago

Δ I guess I never really looked at it as music serving different purposes. I can see how pop music fills a certain purpose, and I guess the artist doesn't necessarily have to be the one to write the song (although I appreciate it when they do).

r/ChangeMyView

OP

CMV: most hit music artists today are bad musicians

1 year ago by * (last edited 1 year ago) [ssbeluga](#)

They're super repetitive, the lyrics and painfully generic, and there's hardly ever anything new or challenging. And from what I understand most of these artists don't even write their own songs. It seems to me they're only popular because of who they are and how they look/perform. Can anyone convince me otherwise?



PC

[–] [kublahkoala](#) 229Δ 13 points 1 year ago

Music appreciation is a skill, and it's all about pattern recognition.

When we're children, we need songs that are really simple, with easy to recognize patterns. The younger we are, the simpler the songs. Toddlers like nursery rhymes. Teens like pop music.

You have to consider that music can be listened to for different purposes. Maybe you need something on in the background during a party, or while you study, or are trying to fall asleep, or work out. Pop music is really good in some of these situations, really bad in others.

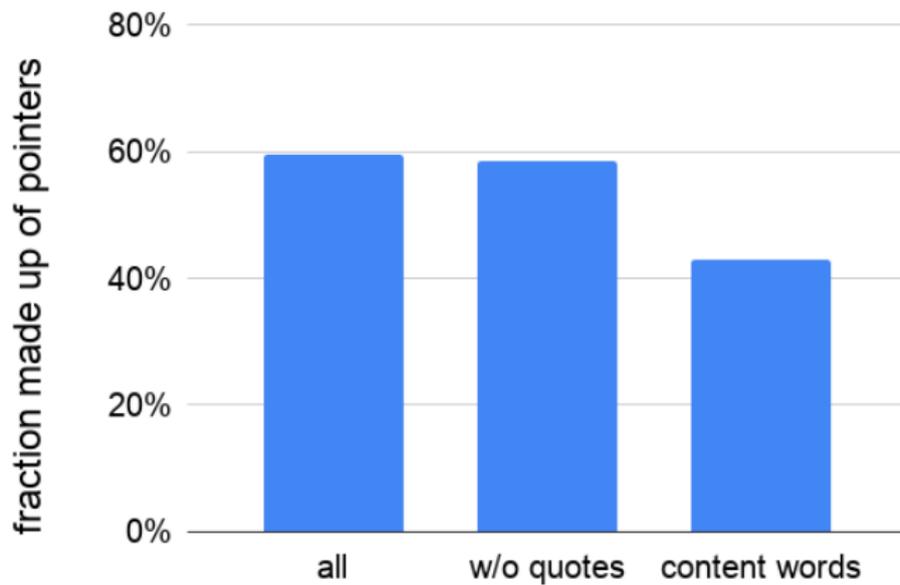
Explanation



[–] [ssbeluga](#) [S] 3 points 1 year ago

Δ I guess I never **really looked** at it as **music serving different purposes**. I can **see** how **pop music** fills a certain **purpose**, and I guess the **artist** doesn't necessarily have to be the **one** to **write** the **song** (although I **appreciate** it when they do).

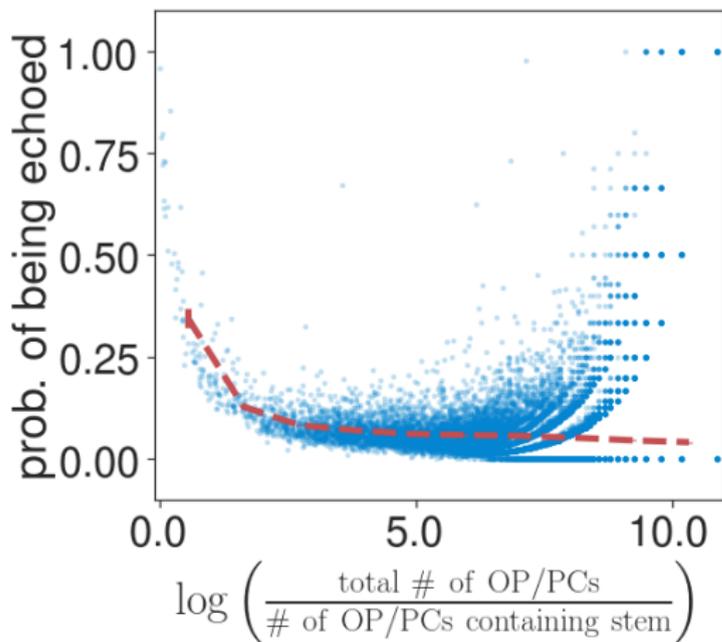
Pointers are common



How do explanations **selectively** incorporate pointers from their explananda?

Probability of echoing vs. word frequency

One answer:



A prediction task!

The task

1. Take the **set of unique stems** in the expandum.
2. For every such stem s , we assign the label

$$\begin{cases} 1 & \text{if } s \text{ is in the set of unique stems in the } \text{explanation} \\ 0 & \text{otherwise.} \end{cases}$$

What *could* affect pointer use?

For example:

1. **Non-contextual properties**
 2. OP and PC usage
 3. How the word connects the OP and PC
 4. General properties of the OP or PC
- IDF (↓)
 - Word length (↓)

What *could* affect pointer use?

1. Non-contextual properties
2. **OP and PC usage**
3. How the word connects the OP and PC
4. General properties of the OP or PC

For example:

- POS tags (verb in OP: ↓, verb in PC: ↑, noun in OP: ↓, noun in PC: ↓).
- Term frequency (↑)
- # of surface forms (↑)
- in a quotation (↑)

What *could* affect pointer use?

1. Non-contextual properties
2. OP and PC usage
3. **How the word connects the OP and PC**
4. General properties of the OP or PC

For example:

- Word is in both OP and PC (↑)
- # of word's surface forms in OP but not in PC (↓), and vice versa (↑)
- JS divergence between OP and PC POS distributions for word (↓)

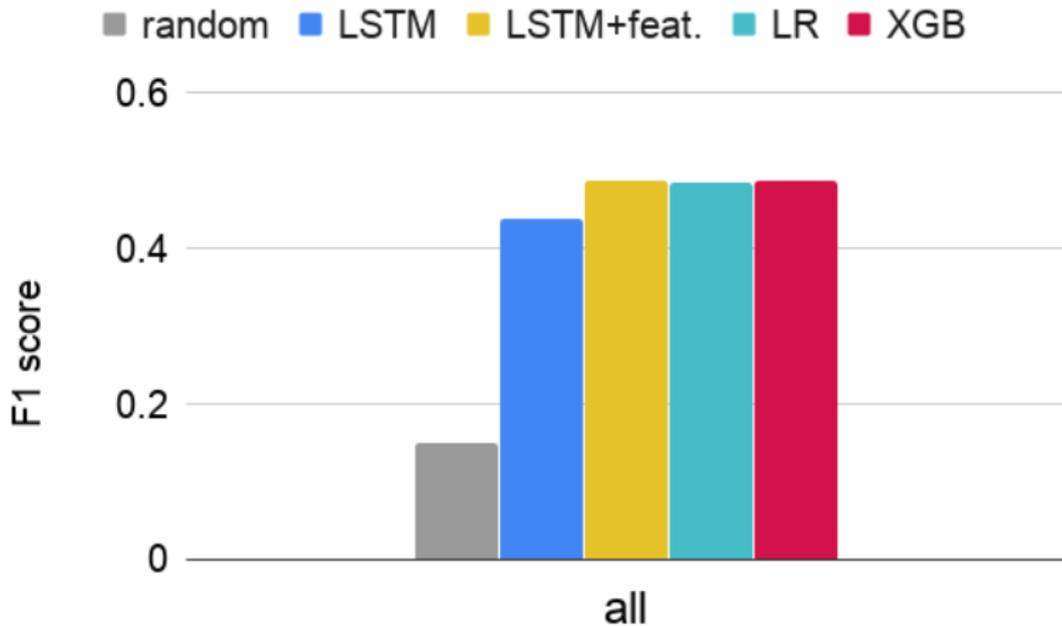
What *could* affect pointer use?

1. Non-contextual properties
2. OP and PC usage
3. How the word connects the OP and PC
4. **General properties of the OP or PC**

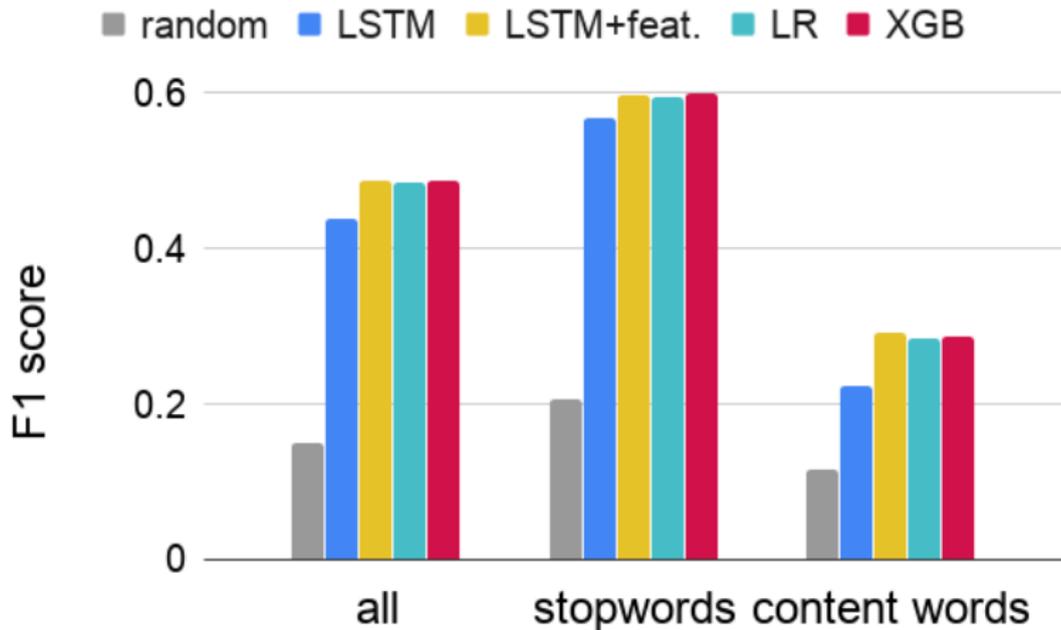
For example:

- OP length (↓), PC length (↑)
- Depth of PC in the thread (↑)
- Difference between the avg. word lengths in OP and PC (↓)

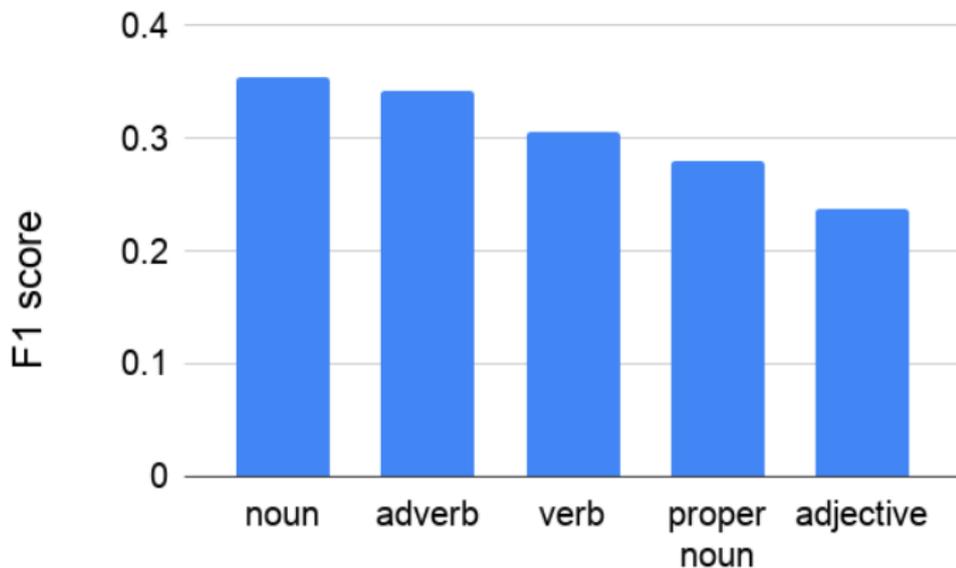
Our features improve on LSTMs



Our features improve on LSTMs

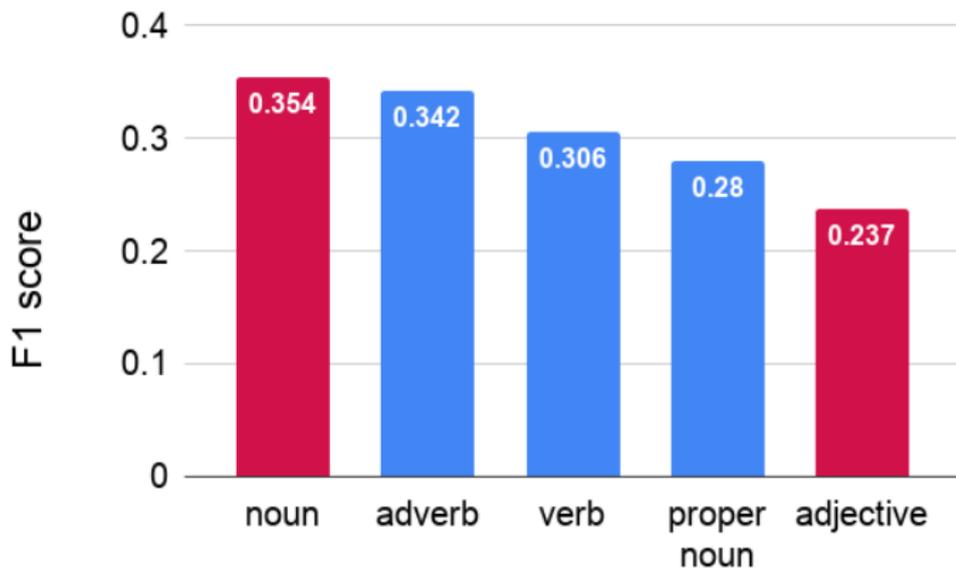


Some parts of speech are more reliably predicted

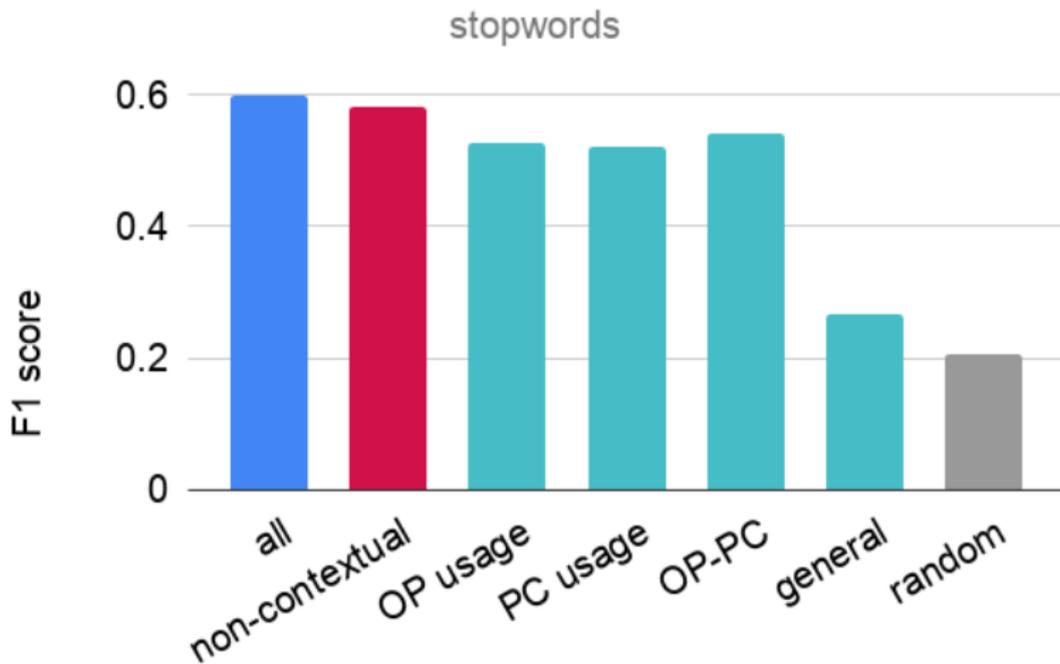


9

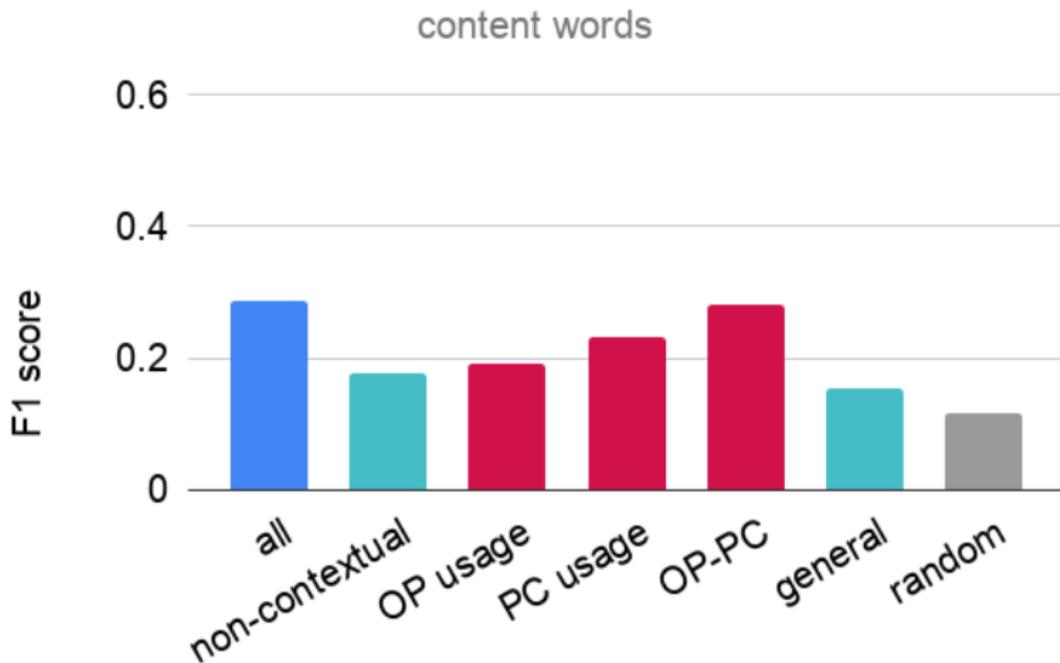
Some parts of speech are more reliably predicted



Which features matter?



Which features matter?



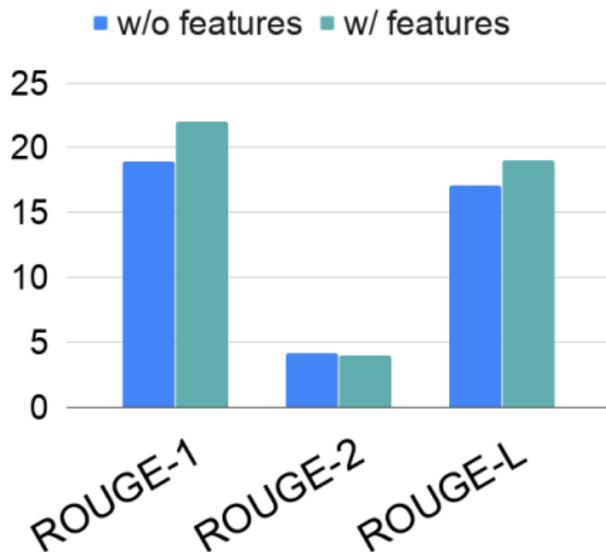
Our features can improve the *generation* of explanations

Pointer generator
network, with
coverage¹⁰
+
our features

¹⁰(See et al., 2017; Klein et al., 2017)

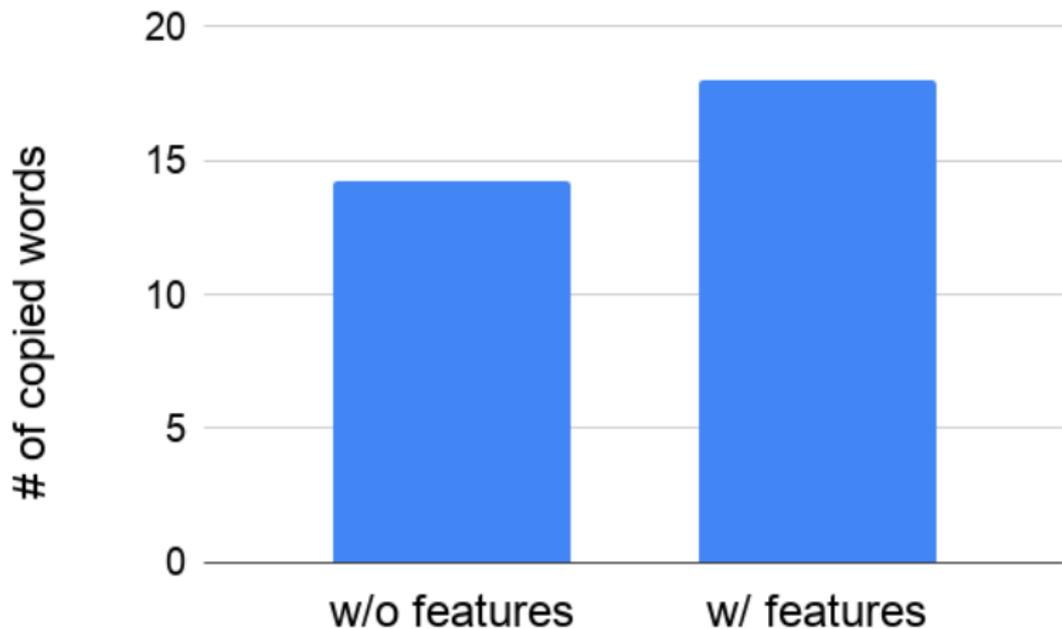
Our features can improve the *generation* of explanations

Pointer generator
network, with
coverage¹⁰
+
our features

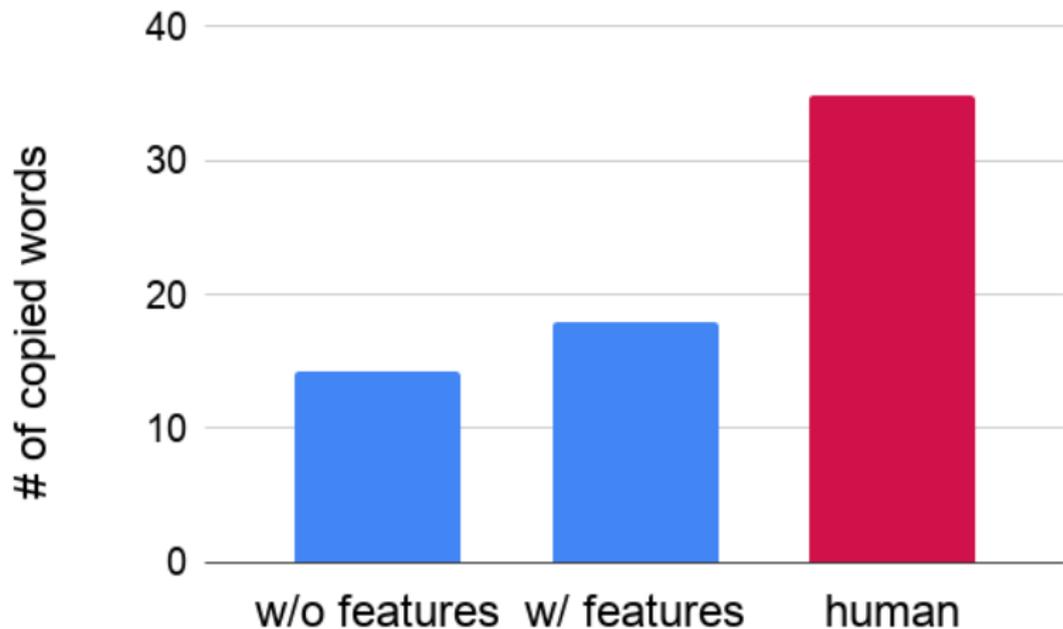


¹⁰(See et al., 2017; Klein et al., 2017)

...and increase copying



...and increase copying



Takeaways

Our Dataset:

- 1. We assemble a novel, large-scale dataset of naturally occurring explanations.**

Takeaways

Our Dataset:

1. We assemble a novel, large-scale dataset of naturally occurring explanations.

Our Findings:

2. **Pointers are common.**

Takeaways

Our Dataset:

1. We assemble a novel, large-scale dataset of naturally occurring explanations.

Our Findings:

2. Pointers are common.
3. **Importance of nouns.**

Takeaways

Our Dataset:

1. We assemble a novel, large-scale dataset of naturally occurring explanations.

Our Findings:

2. Pointers are common.
3. Importance of nouns.
4. **Non-contextual properties for stopwords, contextual for content words.**

Takeaways

Our Dataset:

1. We assemble a novel, large-scale dataset of naturally occurring explanations.

Our Findings:

2. Pointers are common.
3. Importance of nouns.
4. Non-contextual properties for stopwords, contextual for content words.

Our Features:

5. **Improve on prediction performance of vanilla LSTMs.**

Takeaways

Our Dataset:

1. We assemble a novel, large-scale dataset of naturally occurring explanations.

Our Findings:

2. Pointers are common.
3. Importance of nouns.
4. Non-contextual properties for stopwords, contextual for content words.

Our Features:

5. Improve on prediction performance of vanilla LSTMs.
6. **Improve quality of generated explanations.**

Takeaways

Our Dataset:

1. We assemble a novel, large-scale dataset of naturally occurring explanations.

Our Findings:

2. Pointers are common.
3. Importance of nouns.
4. Non-contextual properties for stopwords, contextual for content words.

Our Features:

5. Improve on prediction performance of vanilla LSTMs.
6. Improve quality of generated explanations.

Thank you!

Code + data: github.com/davatk/what-gets-echoed