A Few Open Questions

What is the limit of text-based meaning representations?
Should we be learning this way? Is it data-efficient/effective?
Why should models learn this way, but not humans?
A Few Open Questions

What is the limit of text-based meaning representations?
Should we be learning this way? Is it data-efficient/effective?
Why should models learn this way, but not humans?

NLP’s Answer?
A Few Open Questions

What is the limit of text-based meaning representations?
Should we be learning this way? Is it data-efficient/effective?
Why should models learn this way, but not humans?

NLP’s Answer?
Intelligence and NLU
don’t require seeing,
hearing, or doing, ...
A Few Open Questions

What is the limit of text-based meaning representations?
Should we be learning this way? Is it data-efficient/effective?
Why should models learn this way, but not humans?

NLP’s Answer?

Intelligence and NLU don’t require seeing, hearing, or doing, ...
A Few Open Questions

What is the limit of text-based meaning representations?
Should we be learning this way? Is it data-efficient/effective?
Why should models learn this way, but not humans?

NLP’s Answer?
Intelligence and NLU don’t require seeing, hearing, or doing, …
but might need speaking.
A Few Open Questions

What is the limit of text-based meaning representations?
Should we be learning this way? Is it data-efficient/effective?
Why should models learn this way, but not humans?

NLP’s Answer?

Intelligence and NLU
don’t require seeing,
hearing, or doing, ... but might need speaking.
+158b parameters
+158b parameters
->
8.2% gain on LAMBADA
+158b parameters

->

8.2% gain on LAMBADA

->

423 more correct instances
+158b parameters

->

8.2% gain on LAMBADA

->

423 more correct instances

->

~370 million parameters needed for each new correct instance!
World Scopes

Corpus
World Scopes

Corpus

Internet
World Scopes
World Scopes
World Scopes

- Corpus
- Internet
- Perception
- Embodied
- Social
World Scopes

Corpus  
Internet  
Perception  
Embodied  
Social

How does reporting bias change?
How does reporting bias change?

What new knowledge do models have access to?
World Scopes

How does reporting bias change?
What new knowledge do models have access to?
What advances in modeling and fusion are necessary?
The (linguistic) Internet

The (linguistic) Internet

The only new parameters introduced during from pre-training (Chelba et al. 2013) in order to...
WS3 — How much “knowledge” is in an image?
WS3 — How much “knowledge” is in an image?

Is it every unary relationship?
WS3 — How much “knowledge” is in an image?

Is it every unary relationship?
- Water bottles are clear
- Water bottles are metallic
- Mugs have words
- Water bottles have words
- …
WS3 — How much “knowledge” is in an image?

Is it every unary relationship?
- Water bottles are clear
- Water bottles are metallic
- Mugs have words
- Water bottles have words
- ...

Is it every pairwise relationship?
WS3 — How much “knowledge” is in an image?

Is it every unary relationship?
- Water bottles are clear
- Water bottles are metallic
- Mugs have words
- Water bottles have words
... 

Is it every pairwise relationship?
- Water bottles go on desks,
- Slippers go on floors,
- Poofs are smaller than desks
- Poofs are larger than slippers
...
WS3 — How much “knowledge” is in an image?

Is it every unary relationship?
Water bottles are clear
Water bottles are metallic
Mugs have words
Water bottles have words
...

Is it every pairwise relationship?
Water bottles go on desks,
Slippers go on floors,
Poofs are smaller than desks
Poofs are larger than slippers
...

Is it that the privacy cover on the webcam is still up so I probably just got off a call indicating it’s during the work day and since the water bottle is full and there’s a small plate on the desk, can we infer that Yonatan is taking work meetings during lunch?
WS3 — How much “knowledge” is in an image?

Is it every unary relationship?
- Water bottles are clear
- Water bottles are metallic
- Mugs have words
- Water bottles have words
  ...

Is it every pairwise relationship?
- Water bottles go on desks,
- Slippers go on floors,
- Poofs are smaller than desks
- Poofs are larger than slippers
  ...

Is it that the privacy cover on the webcam is still up so I probably just got off a call indicating it’s during the work day and since the water bottle is full and there’s a small plate on the desk, can we infer that Yonatan is taking work meetings during lunch?

Yes
WS3 - Language beyond Text

Gestures
WS3 - Language beyond Text

Gestures
WS3 - Language beyond Text

Gestures     Facial expression
WS3 - Language beyond Text

Gestures  Facial expression
WS3 - Language beyond Text

Gestures   Facial expression

(Even just emojis 😃)
WS3 - Language beyond Text

- Gestures
- Facial expression
- Intonation/Stress

(Even just emojis 😃)
I didn't take the test yesterday.
(Somebody else did.)
I didn't take the test yesterday.
(I did not take it.)
I didn't take the test yesterday.
(I did something else with it.)
I didn't take the test yesterday.
I didn't take the test yesterday.
I didn't take the test yesterday.

WS3 - Language beyond Text

Gestures | Facial expression | Intonation/Stress
---|---|---
(Even just emojis 😊) | | (eVeN JuSt fOnTs)

I didn't take the test yesterday.
(Somebody else did.)
I *didn't* take the test yesterday.
(I did *not* take it.)
I didn't *take* the test yesterday.
(I did something else with it.)
I didn't take the *test* yesterday.
I didn't take the test *yesterday*.

WS4 — How much “knowledge” is in an environment?

https://www.youtube.com/watch?v=HgjpMtRpNrM
WS4 — How much “knowledge” is in an environment?

https://www.youtube.com/watch?v=HgjpMtRpNrM
WS4 — How much “knowledge” is in an environment?

Is it every object affordance?

https://www.youtube.com/watch?v=HsqaMBlshA4
WS4 — How much “knowledge” is in an environment?

Is it every object affordance?

Blocks can stack.
Well, some blocks can stack.
Wheels can roll.
Wheels experience friction.
...

https://www.youtube.com/watch?v=HgjpMrRlQdM
WS4 — How much “knowledge” is in an environment?

Is it every object affordance?

Blocks can stack.
Well, some blocks can stack.
Wheels can roll.
Wheels experience friction.

Is it every possible state or transition?
WS4 — How much “knowledge” is in an environment?

Is it every object affordance?
Blocks can stack.
Well, some blocks can stack.
Wheels can roll.
Wheels experience friction.
...

Is it every possible state or transition?
Car can have 3 pink block on top.
Car can have one green on top.
Car can slide forward or backward.
Top pink block can be removed.
...

https://www.youtube.com/watch?v=HgjpMtRpNrM
WS4 — How much “knowledge” is in an environment?

Is it every object affordance?
Blocks can stack.
Well, some blocks can stack.
Wheels can roll.
Wheels experience friction.
...

Is it every possible state or transition?
Car can have 3 pink block on top.
Car can have one green on top.
Car can slide forward or backward.
Top pink block can be removed.
...

It starts to seem silly to even imagining using language to enumerate the possibilities of an environment, even a tabletop with blocks. The affordances, states, and transitions we learn for planning are strikingly mundane to express verbally, even on something like WikiHow.
WS4 - The Language of the Physical World
WS4 - The Language of the Physical World

Object Representations
WS4 - The Language of the Physical World

Object Representations
WS4 - The Language of the Physical World

Object Representations

- Banana
- Orange
- Baseball
- Toy Castle
WS4 - The Language of the Physical World

Object Representations

Physical Reasoning
WS4 - The Language of the Physical World

Object Representations

Physical Reasoning

?
WS4 - The Language of the Physical World

Object Representations

Physical Reasoning

[Image of objects and reasoning process]

[Bakhtin et al., NeurIPS'19]
WS4 - The Language of the Physical World

Object Representations

Physical Reasoning

A distant concern
A real moonshot
Thick tension
Heavy emotions
WS4 - The Language of the Physical World

Object Representations

Physical Reasoning

Affordances and Plans

A distant concern
A real moonshot
Thick tension
Heavy emotions
WS4 - The Language of the Physical World

Object Representations

Physical Reasoning

Affordances and Plans

A distant concern
A real moonshot
Thick tension
Heavy emotions
WS4 - The Language of the Physical World

Object Representations

Physical Reasoning

Affordances and Plans

A distant concern
A real moonshot
Thick tension
Heavy emotions

[Shridhar et al., CVPR'20]
WS4 - The Language of the Physical World

Object Representations

Physical Reasoning

Affordances and Plans

A distant concern
A real moonshot
Thick tension
Heavy emotions

Should I hammer a nail with
1) a screwdriver or
2) a rock?

[Bisk et al., AAAI’20]
annoying person: am I annoying?

everyone: *annoyed* no
What delimits a World Scope?
What delimits a World Scope?

You can’t learn language ...
What delimits a World Scope?

You can’t learn language ... ...

... from the radio (Internet).  \( WS2 \subset WS3 \)

A task learner cannot be said to be in WS3 if it can succeed without perception (e.g., visual, auditory).
What delimits a World Scope?

You can’t learn language ... 

... from the radio (Internet).  $\text{WS2} \subset \text{WS3}$

A task learner cannot be said to be in WS3 if it can succeed without perception (e.g., visual, auditory).

... from a television.  $\text{WS3} \subset \text{WS4}$

A task learner cannot be said to be in WS4 if the space of its world actions and consequences can be enumerated.
What delimits a World Scope?

You can’t learn language ...

... from the radio (Internet). \( \text{WS2} \subset \text{WS3} \)

A task learner cannot be said to be in WS3 if it can succeed without perception (e.g., visual, auditory).

... from a television. \( \text{WS3} \subset \text{WS4} \)

A task learner cannot be said to be in WS4 if the space of its world actions and consequences can be enumerated.

... by yourself. \( \text{WS4} \subset \text{WS5} \)

A task learner cannot be said to be in WS5 unless achieving its goals requires cooperating with a human in the loop.
What should we do about this?
Rethink every task
Rethink every task

Include phenomena from WS3-5 in your model’s universe …
Rethink every task

Include phenomena from WS3-5 in your model’s universe …

Will grounding make generation more controllable?  Yes.
Rethink every task

Include phenomena from WS3-5 in your model’s universe …

Will grounding make generation more controllable? Yes.
Should model evaluation use humans-in-the-loop? Yes.
Rethink every task

Include phenomena from WS3-5 in your model’s universe …

Will grounding make generation more controllable? Yes.
Should model evaluation use humans-in-the-loop? Yes.
Can MT systems use interaction for more targeted learning? Yes.
Rethink every task

Include phenomena from WS3-5 in your model’s universe …

Will grounding make generation more controllable? Yes.
Should model evaluation use humans-in-the-loop? Yes.
Can MT systems use interaction for more targeted learning? Yes.
Can physical exploration improve knowledge of entailment? Yes.
Rethink every task

Include phenomena from WS3-5 in your model’s universe …

Will grounding make generation more controllable? Yes.
Should model evaluation use humans-in-the-loop? Yes.
Can MT systems use interaction for more targeted learning? Yes.
Can physical exploration improve knowledge of entailment? Yes.
Can perception resolve syntactic/anaphoric/discursive ambiguity? Yes.
Rethink every task

Include phenomena from WS3-5 in your model’s universe …

Will grounding make generation more controllable? Yes.
Should model evaluation use humans-in-the-loop? Yes.
Can MT systems use interaction for more targeted learning? Yes.
Can physical exploration improve knowledge of entailment? Yes.
Can perception resolve syntactic/anaphoric/discursive ambiguity? Yes.
Can tests in a simulator capture distinctions that MC struggles with? Yes.
Rethink every task

Include phenomena from WS3-5 in your model’s universe …

Will grounding make generation more controllable? Yes.
Should model evaluation use humans-in-the-loop? Yes.
Can MT systems use interaction for more targeted learning? Yes.
Can physical exploration improve knowledge of entailment? Yes.
Can perception resolve syntactic/anaphoric/discursive ambiguity? Yes.
Can tests in a simulator capture distinctions that MC struggles with? Yes.

…and consider where signal for your task comes from!
Is knowledge about task $X$ most richly encoded in:
Is knowledge about task X most richly encoded in:
Is knowledge about task X most richly encoded in: