Do contextualised word-embeddings represent richly subsective adjectives more diversely?

Michael Goodale and Salvador Mascarenhas
Institut Jean Nicod — École normale supérieure

August 10, 2022

Michael Goodale and Salvador Mascarenhas Institut Jean Nicod — École normale supérieure

Do contextualised word-embeddings represent richly subsective adjectives more diversely?

Distributional representations and adjectives

- A key part of language models in NLP are the representations of words.
- Generally, representations were *static*, each word has one representation in all contexts.
- These static representations capture a lot of semantic and syntactic information, but clearly struggle with even basic homonymy/polysemy.
- A systematic kind of polysemy occurs with certain kinds of adjectives.

Michael Goodale and Salvador Mascarenhas Institut Jean Nicod — École normale supérieure

Do contextualised word-embeddings represent richly subsective adjectives more diversely?

Adjectives and word embeddings

A classical typology of adjectives divides them into three groups (sometimes four):

Intersective	Subsective	Privative/Intensional
French, bald, dead	skilled, good	fake, alleged

Michael Goodale and Salvador Mascarenhas   Institut Jean Nicod — École normale supérieure

Do contextualised word-embeddings represent richly subsective adjectives more diversely?

Adjectives and word-embeddings
○○●○○○○○○○

Study
○○○○○

Results
○○○○○○○

References

Intersective Adjectives

The differences with these adjectives lies with the entailments one can make:

Michael Goodale and Salvador Mascarenhas Institut Jean Nicod — École normale supérieure

Do contextualised word-embeddings represent richly subsective adjectives more diversely?

Intersective Adjectives

The differences with these adjectives lies with the entailments one can make:

> He is **French** programmer.
> He is a writer.
> ∴   He is a **French** writer.   ✓

Michael Goodale and Salvador Mascarenhas Institut Jean Nicod — École normale supérieure

Do contextualised word-embeddings represent richly subsective adjectives more diversely?

Subsective Adjectives

Subsectives are more specific to the noun they modify:

He is a **good** programmer.
He is a writer.

∴ He is a **good** writer.   ✗

Michael Goodale and Salvador Mascarenhas Institut Jean Nicod — École normale supérieure

Do contextualised word-embeddings represent richly subsective adjectives more diversely?

Rich and weak subsective adjectives

Subsectivity has multiple origins:

1. A different standard degree.
2. A broader difference in meaning.

He is a **short** basket-ball player.
∴   He is a **short** person. ✗

(a) An *weakly-subsective* adjective:
$[\![$short person$]\!] =$
$\lambda x.person(x) \wedge [\![$short-for-$a]\!](x)$

He is a **horrible** singer.
∴   He is a **horrible** person. ✗

(b) A *richly-subsective* adjective:
$[\![$horrible person$]\!] =$
$\lambda x.person(x) \wedge [\![$horrible-at-$a$-for-$b]\!](x)$

Basic idea: Richly-subsective adjectives require a richer free variable than weakly subsective-adjectives need only a degree.

Michael Goodale and Salvador Mascarenhas Institut Jean Nicod — École normale supérieure

Do contextualised word-embeddings represent richly subsective adjectives more diversely?

Adjectives and word-embeddings
○○○○○○●○○○

Study
○○○○○

Results
○○○○○○○

References

Contextuality and subsectivity

(1) Last night on Dancing with the Stars, there was that really talented actor, but also that terrible scientist.

Michael Goodale and Salvador Mascarenhas Institut Jean Nicod — École normale supérieure

Do contextualised word-embeddings represent richly subsective adjectives more diversely?

Adjectives and word-embeddings
○○○○○○○●○○○

Study
○○○○○

Results
○○○○○○○

References

Intensional adjectives

A final class of adjectives go a step further and change entailments about the noun itself.

He is **fake** doctor.

He is a writer.

∴   He is a **fake** writer. ✗

∴   He is a doctor. ✗

Michael Goodale and Salvador Mascarenhas Institut Jean Nicod — École normale supérieure

Do contextualised word-embeddings represent richly subsective adjectives more diversely?

Previous research

Some work has previously looked to see if distributional methods are sensitive to these distinctions (Boleda et al. 2013).

Basic idea was comparing different ways of combining adjective and noun vectors to produce vectors for an adjective+noun phrase compositionally (Baroni and Zamparelli 2010).

Michael Goodale and Salvador Mascarenhas Institut Jean Nicod — École normale supérieure

Do contextualised word-embeddings represent richly subsective adjectives more diversely?

Adjectives and word-embeddings
○○○○○○○○○●○

Study
○○○○○

Results
○○○○○○○

References

Previous research

No difference between intensional adjectives and non-intensional adjectives
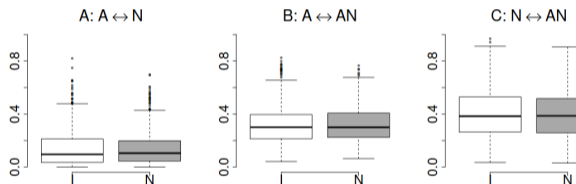

Figure 1: Distribution of cosines for observed vectors, by adjective type (intensional, I, or non-intensional, N). From left to right, adjective vs. noun, adjective vs. phrase, and noun vs. phrase cosines.

Michael Goodale and Salvador Mascarenhas Institut Jean Nicod — École normale supérieure

Do contextualised word-embeddings represent richly subsective adjectives more diversely?

Previous research

No difference between intensional adjectives and non-intensional adjectives

Figure 1: Distribution of cosines for observed vectors, by adjective type (intensional, I, or non-intensional, N). From left to right, adjective vs. noun, adjective vs. phrase, and noun vs. phrase cosines.

We look at *contextual* word-embeddings to see if they are sensitive to these distinctions.

Michael Goodale and Salvador Mascarenhas Institut Jean Nicod — École normale supérieure

Do contextualised word-embeddings represent richly subsective adjectives more diversely?

Adjectives and word-embeddings
○○○○○○○○○●

Study
○○○○○

Results
○○○○○○○

References

Contextual word embeddings

In 2018 models like ELMo introduced *contextual* word embeddings which are now used basically by all language models. Here, words get *different* representations for each sentence.
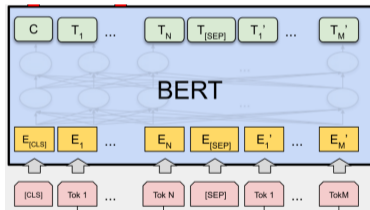
This can better handle homonym pair like river *bank* and financial *bank* or more subtle kinds of polysemy. Importantly, it allows for *sentence-specific* syntactic and semantic information to be encoded.

Michael Goodale and Salvador Mascarenhas Institut Jean Nicod — École normale supérieure

Do contextualised word-embeddings represent richly subsective adjectives more diversely?

Contextual word embeddings and adjectives

Does a richer free variable lead to a more diverse contextual word embedding?

Michael Goodale and Salvador Mascarenhas Institut Jean Nicod — École normale supérieure

Do contextualised word-embeddings represent richly subsective adjectives more diversely?

Adjectives and word-embeddings
○○○○○○○○○○

Study
●○○○○

Results
○○○○○○○

References

Contextual word embeddings and adjectives

Does a richer free variable lead to a more diverse contextual word embedding?

In other words, does "skilled" change between sentences more than "Canadian" for contextual word embeddings?

Michael Goodale and Salvador Mascarenhas Institut Jean Nicod — École normale supérieure

Do contextualised word-embeddings represent richly subsective adjectives more diversely?

Adjectives and word-embeddings
○○○○○○○○○○

Study
○●○○○○

Results
○○○○○○○

References

Methodology

We simply compare how similar adjective representations are across different sentences to see how much they change. To measure similarity, we just take the cosine distance between two different representations of the same adjective modifying different nouns:

Adjectives and word-embeddings
○○○○○○○○○○

Study
○●○○○○

Results
○○○○○○○

References

Methodology

We simply compare how similar adjective representations are across different sentences to see how much they change. To measure similarity, we just take the cosine distance between two different representations of the same adjective modifying different nouns:
Problem: Each contextual word embeddings contains influence from the entire sentence.

Methodology

We simply compare how similar adjective representations are across different sentences to see how much they change. To measure similarity, we just take the cosine distance between two different representations of the same adjective modifying different nouns:
Problem: Each contextual word embeddings contains influence from the entire sentence.
Bigger problem: Words might be used in more diverse contexts independent of the richness of its free variable.

Michael Goodale and Salvador Mascarenhas Institut Jean Nicod — École normale supérieure

Do contextualised word-embeddings represent richly subsective adjectives more diversely?

Solution: "Minimal pairs" and generated sentences

- This is my friend, {name} who is a {adjective} {noun}.
- So, I met someone, {name} is a {adjective} {noun}.
- By the way, {name} is a {adjective} {noun}.
- Did you know that {name} is a {adjective} {noun}?
- In fact, {name} is a {adjective} {noun}.
- By the way, {name} is a {noun} who is {adjective}.

Adjectives

Rich Subsective	good, bad, skilled, typical, talented, normal, terrible, fine, great, horrific, horrible, inferior, dreadful
Weak Subsective	large, fat, nervous, kind, cruel, tall, short, happy, sad, beautiful, ugly, healthy, rich
Intersective	Canadian, Russian, Chinese, Brazilian, white, Black, straight, gay, Christian, Muslim, Jewish, brunette, blond, bald, naked
Intensional	alleged, future, potential, presumed

Michael Goodale and Salvador Mascarenhas Institut Jean Nicod — École normale supérieure

Do contextualised word-embeddings represent richly subsective adjectives more diversely?

Nouns

- doctor
- surgeon
- nurse
- driver
- chef
- waiter

- scientist
- pianist
- gardener
- carpenter
- technician
- athlete

- plumber
- actor
- author
- novelist
- musician
- soldier

- spy
- poet
- person
- runner
- dancer
- architect

Adjectives and word-embeddings
○○○○○○○○○○

Study
○○○○○

Results
●○○○○○○

References

Similarity correlates with "richness" of the variable

(a) `bert-large-uncased`

(b) `sentence-roberta-large`

Michael Goodale and Salvador Mascarenhas Institut Jean Nicod — École normale supérieure

Do contextualised word-embeddings represent richly subsective adjectives more diversely?
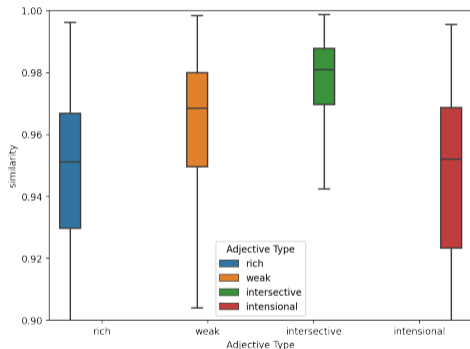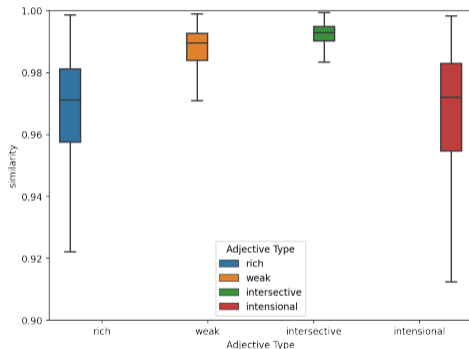
If we change the name and not the modified noun, the pattern disappears

(a) `bert-large-uncased`

(b) `sentence-roberta-large`

Michael Goodale and Salvador Mascarenhas Institut Jean Nicod — École normale supérieure

Do contextualised word-embeddings represent richly subsective adjectives more diversely?
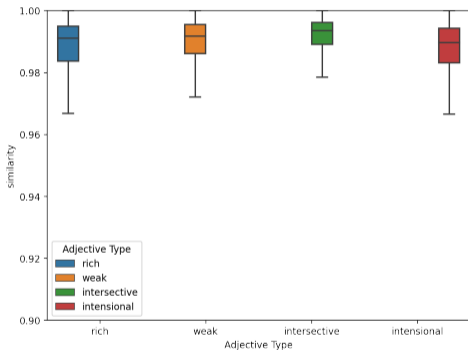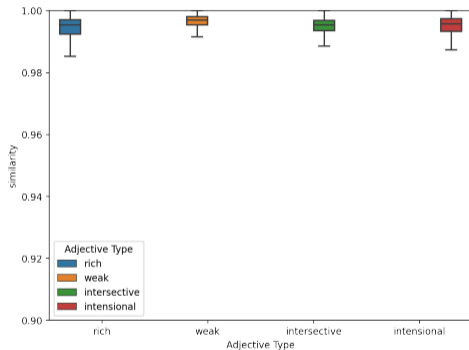
Adjectives vary considerably individually

Michael Goodale and Salvador Mascarenhas Institut Jean Nicod — École normale supérieure

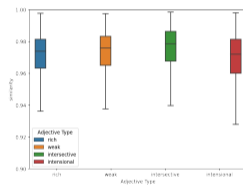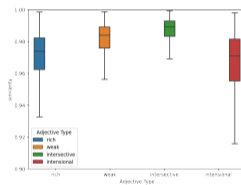Do contextualised word-embeddings represent richly subsective adjectives more diversely?
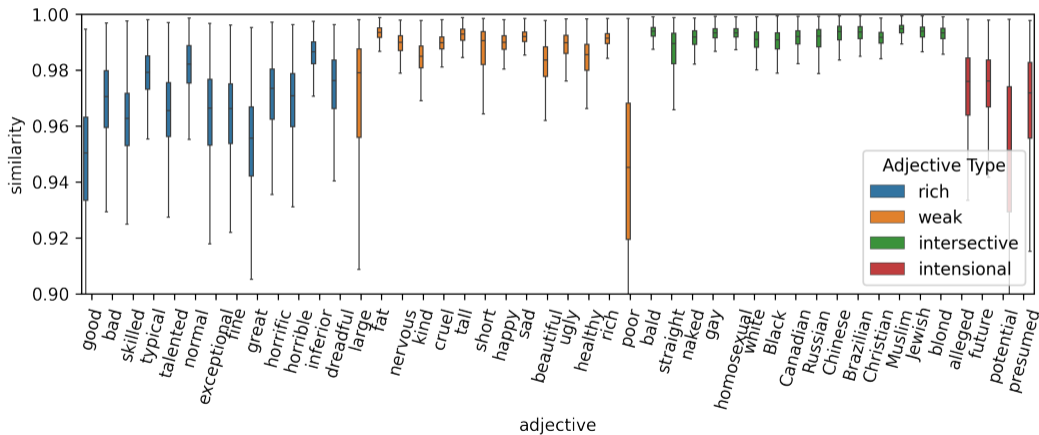
Pattern only emerges in the highest layers of the network
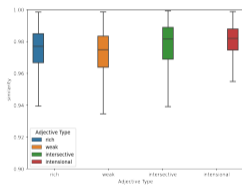
(a) Layer 6   (b) Layer 12   (c) Layer 16   (d) Layer 24

Michael Goodale and Salvador Mascarenhas Institut Jean Nicod — École normale supérieure

Do contextualised word-embeddings represent richly subsective adjectives more diversely?
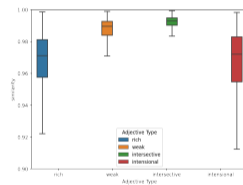
Interpretation

Possibly evidence of "compositionality", the representations seem to change systematically based on the noun they modify if they have a rich free-variable.

It's not wholly surprising that these distinctions are drawn:

(2)   John is a good musician but not a good friend.

(3)   John is a fake musician but not a fake friend.

(4) # John is a tall musician but not a tall friend.

(5) # John is a bald musician but not a bald friend.

Michael Goodale and Salvador Mascarenhas   Institut Jean Nicod — École normale supérieure

Do contextualised word-embeddings represent richly subsective adjectives more diversely?

Next steps

- Larger corpus on more diverse set of objects.
- Train some kind of compositional model to predict final adjective embeddings given the static noun and adjective pairs à la Baroni and Zamparelli 2010 and Boleda et al. 2013.

Michael Goodale and Salvador Mascarenhas Institut Jean Nicod — École normale supérieure

Do contextualised word-embeddings represent richly subsective adjectives more diversely?

Thank you!

References I

[1] Marco Baroni and Roberto Zamparelli. "Nouns are Vectors, Adjectives are Matrices: Representing Adjective-Noun Constructions in Semantic Space". In: *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*. EMNLP 2010. Cambridge, MA: Association for Computational Linguistics, Oct. 2010, pp. 1183–1193. URL: https://aclanthology.org/D10-1115 (visited on 08/10/2022).

[2] Gemma Boleda et al. "Intensionality was only alleged: On adjective-noun composition in distributional semantics". In: *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013) – Long Papers*. Potsdam, Germany: Association for Computational Linguistics, Mar. 2013, pp. 35–46. URL: https://aclanthology.org/W13-0104 (visited on 11/25/2021).

Michael Goodale and Salvador Mascarenhas Institut Jean Nicod — École normale supérieure

Do contextualised word-embeddings represent richly subsective adjectives more diversely?