Situating Sentence Embedders with Nearest Neighbor Overlap

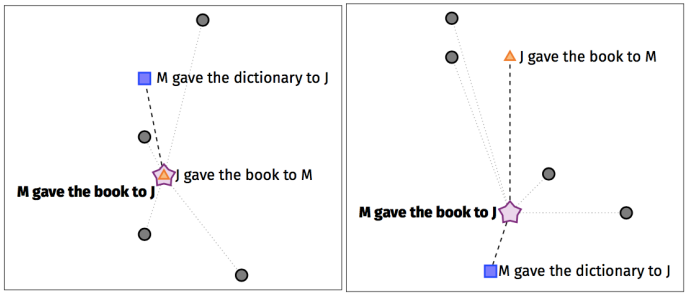
Lucy H. Lin, Noah A. Smith
University of Washington

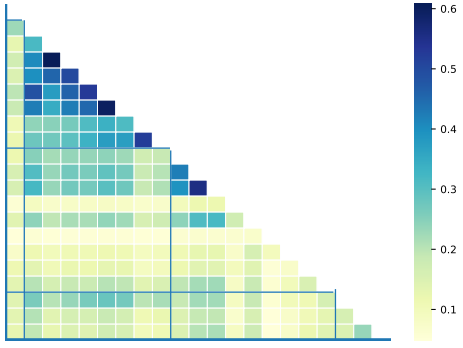
4 October 2019

N2O: a procedure for comparing sentence embedding models

general experimental results

case study: BERT





Query.
Britain's biggest mortgage lender says that average house prices fell 3.6 percent in September, but analysts believe the market isn't that weak.

Near neighbor for `BERT-cls` *only.*
Japanese consumer prices fell for 13th straight month in March, though the GDP data suggests that deflationary pressures are starting to ease.

Text as data

How do we represent text numerically?

- counting n-grams
- word embeddings
- ...
- something via deep learning with big data

Sentences are useful!

"*Federal* government should provide *economic support* to help our *communities* thrive."

"This *federal assistance* would help provide *economic development* in the *community*."

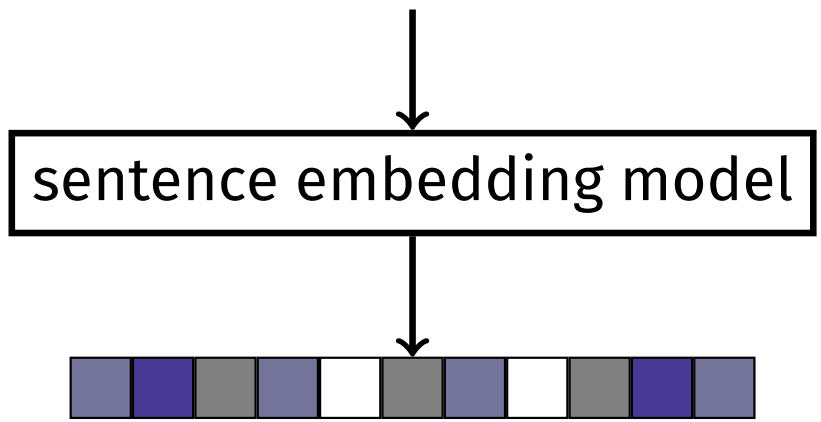
...how do we represent sentences?

(examples from TADA '18: Dreier, Lin, Serrano, Gade, Smith)

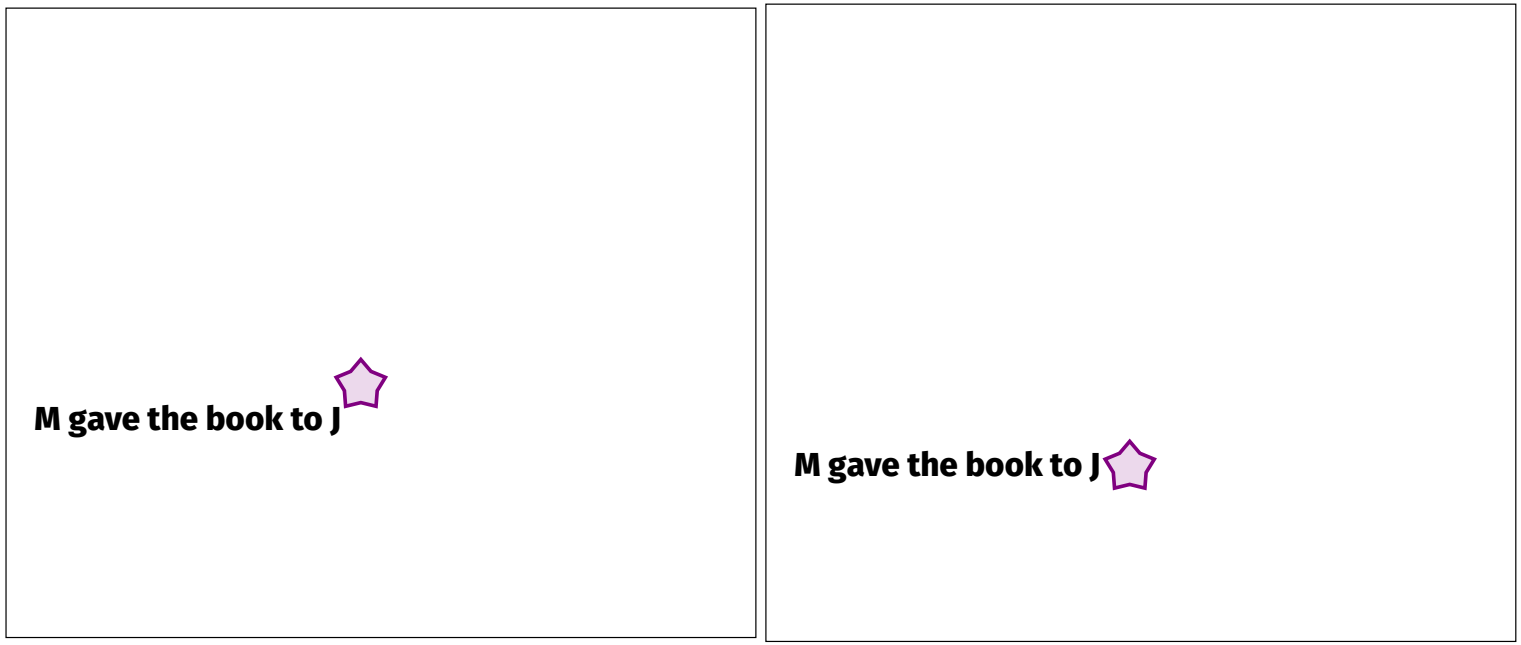
Sentence embedding models

Def. map sentences to fixed-length vectors

Federal government should provide economic support to help our communities thrive.

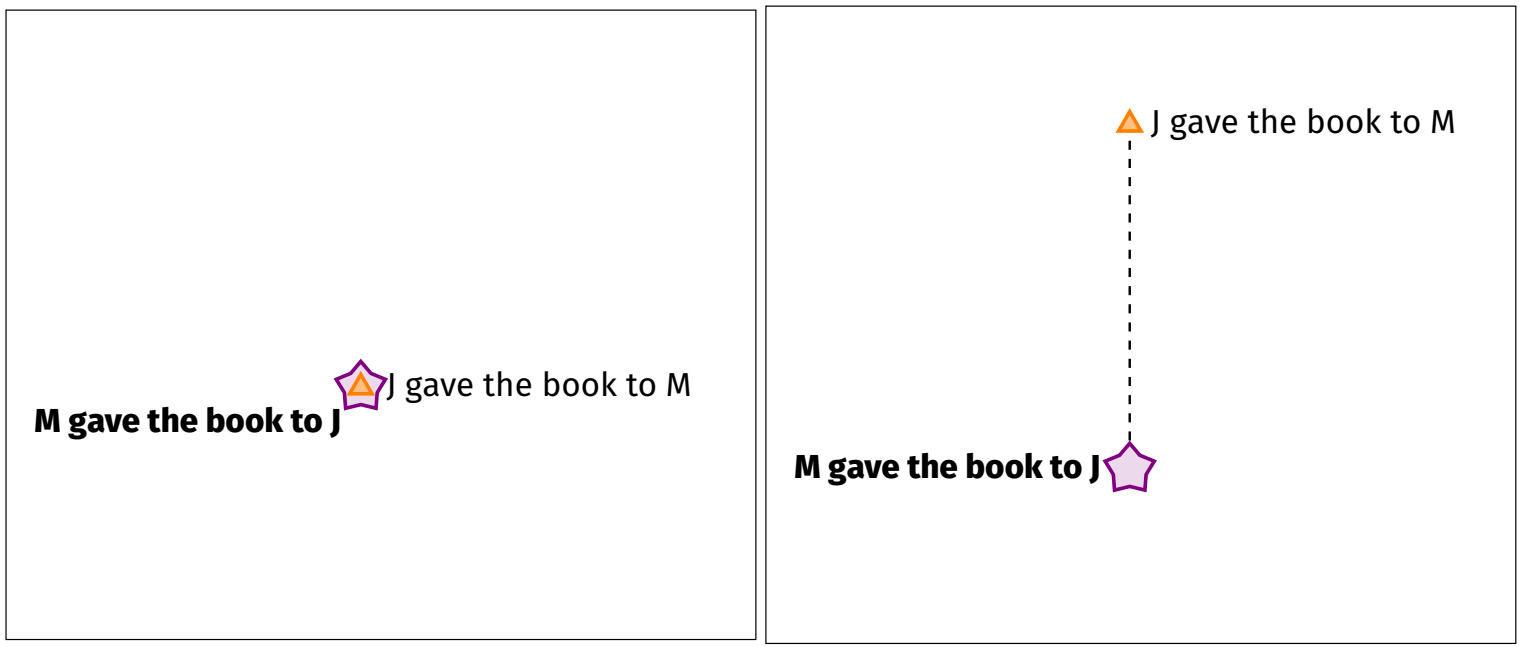


Sentence embedding models



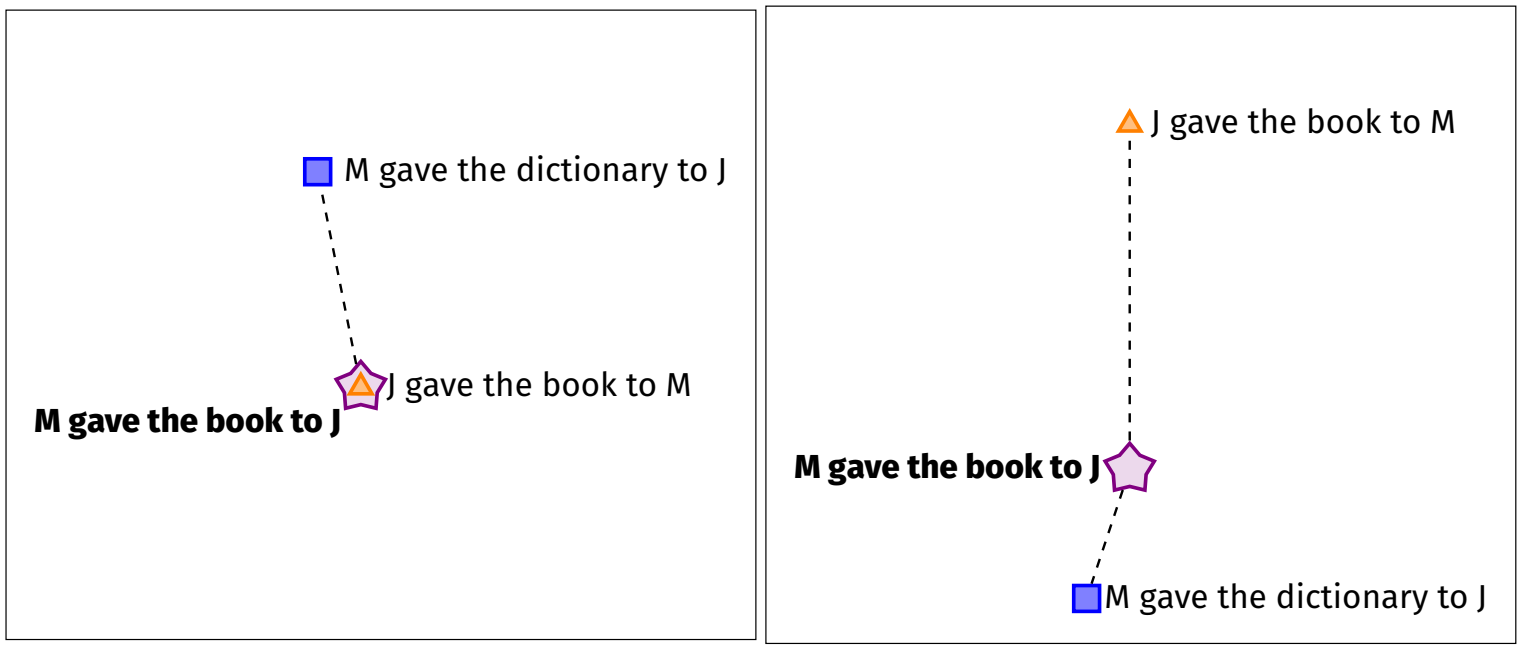
- Different embedding models may put different sentences near each other

Sentence embedding models



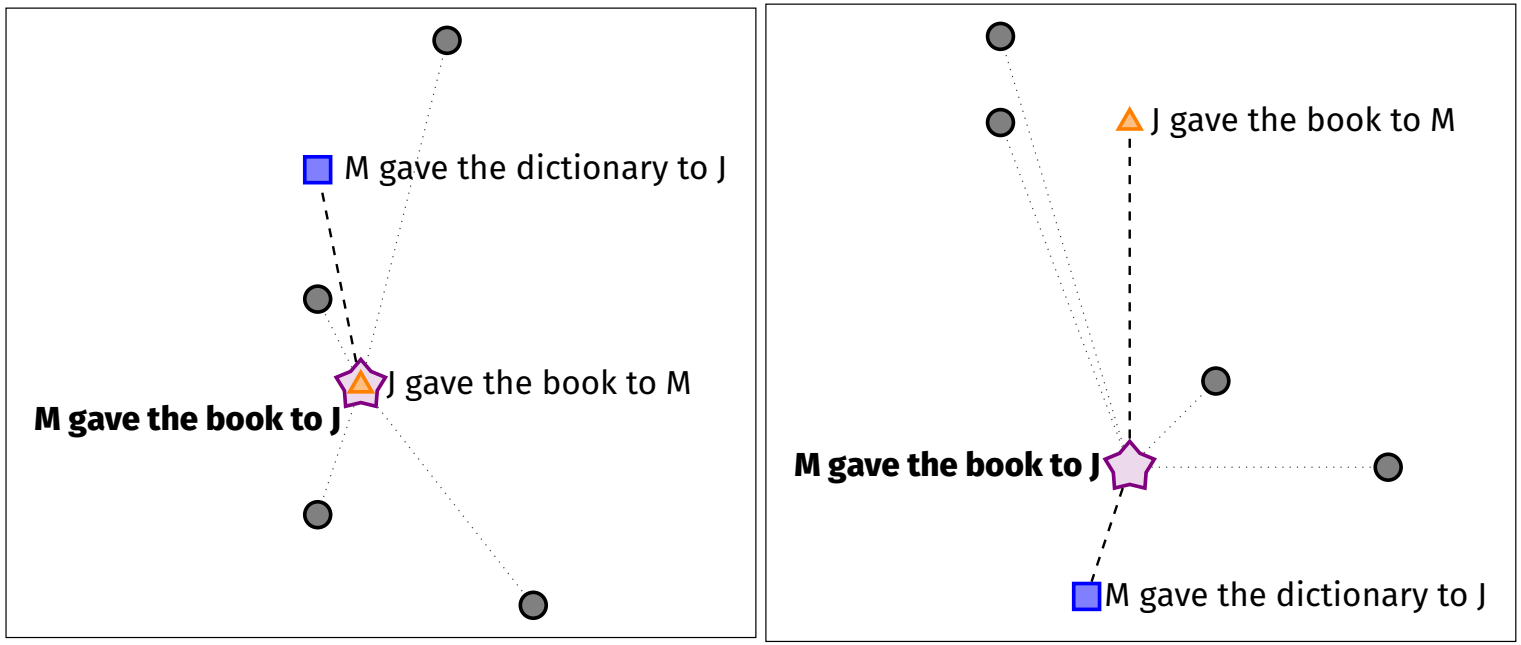
- Different embedding models may put different sentences near each other

Sentence embedding models



- Different embedding models may put different sentences near each other

Sentence embedding models



- Different embedding models may put different sentences near each other

How do we choose?

Extrinsic. downstream task, optimized metrics
(e.g., GLUE benchmark by Wang et al., 2019)

Intrinsic. linguistic/interpretable properties
(Ettinger et al., 2016; Conneau et al., 2018;
Zhu et al., 2018; among many others)

N2O: comparative approach; no additional annotation.

Nearest neighbor overlap (N2O)

Two embedding models are **more similar** if,

for the same set of inputs,

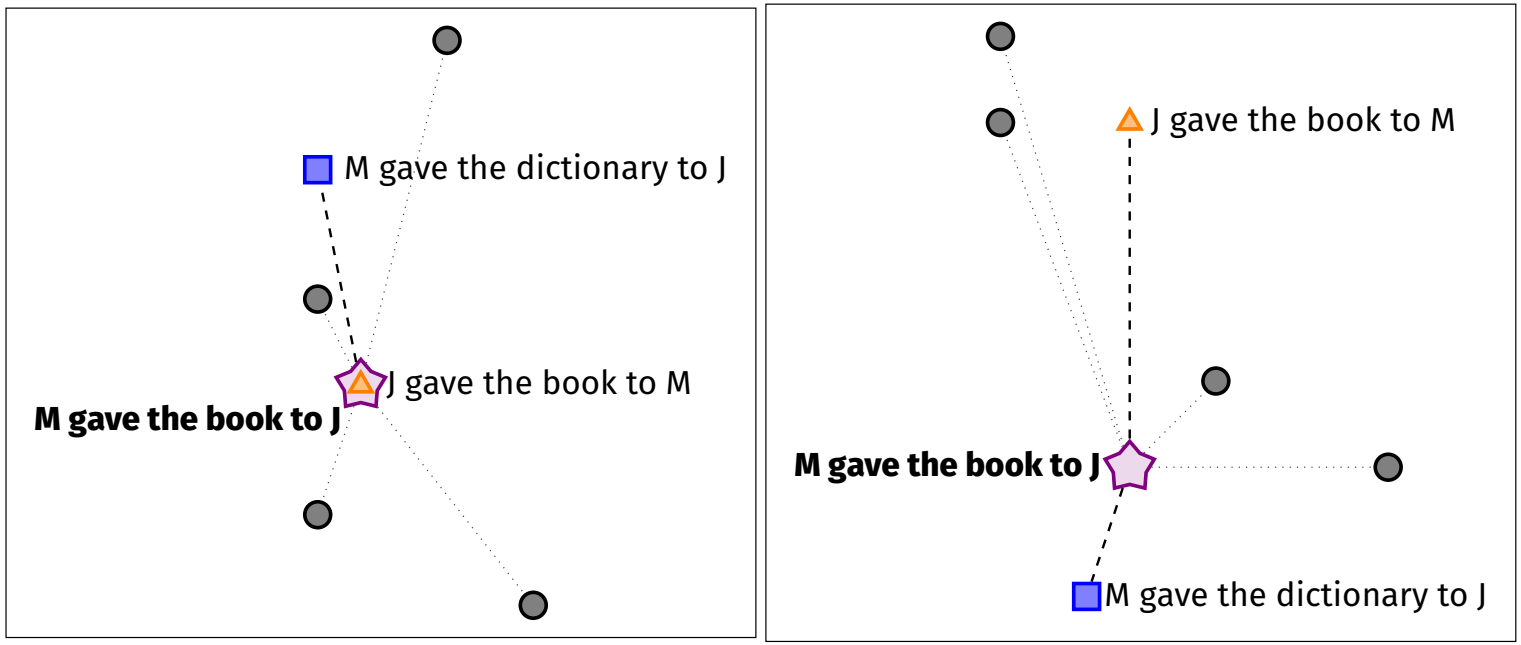
there is **greater overlap** between the

inputs' nearest neighbors.

Common intuition:

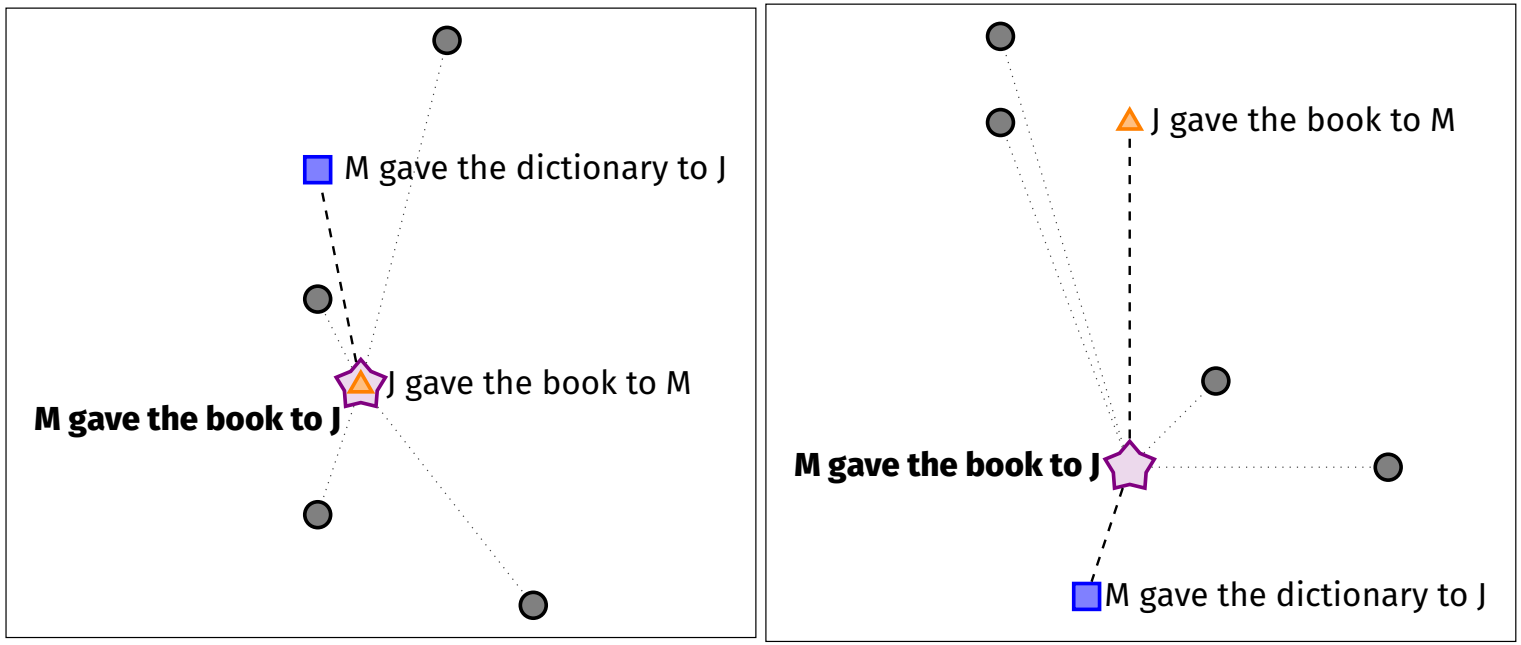
Semantic similarity $\Leftrightarrow$ nearness in embedding space.

Nearest neighbor overlap (N2O)



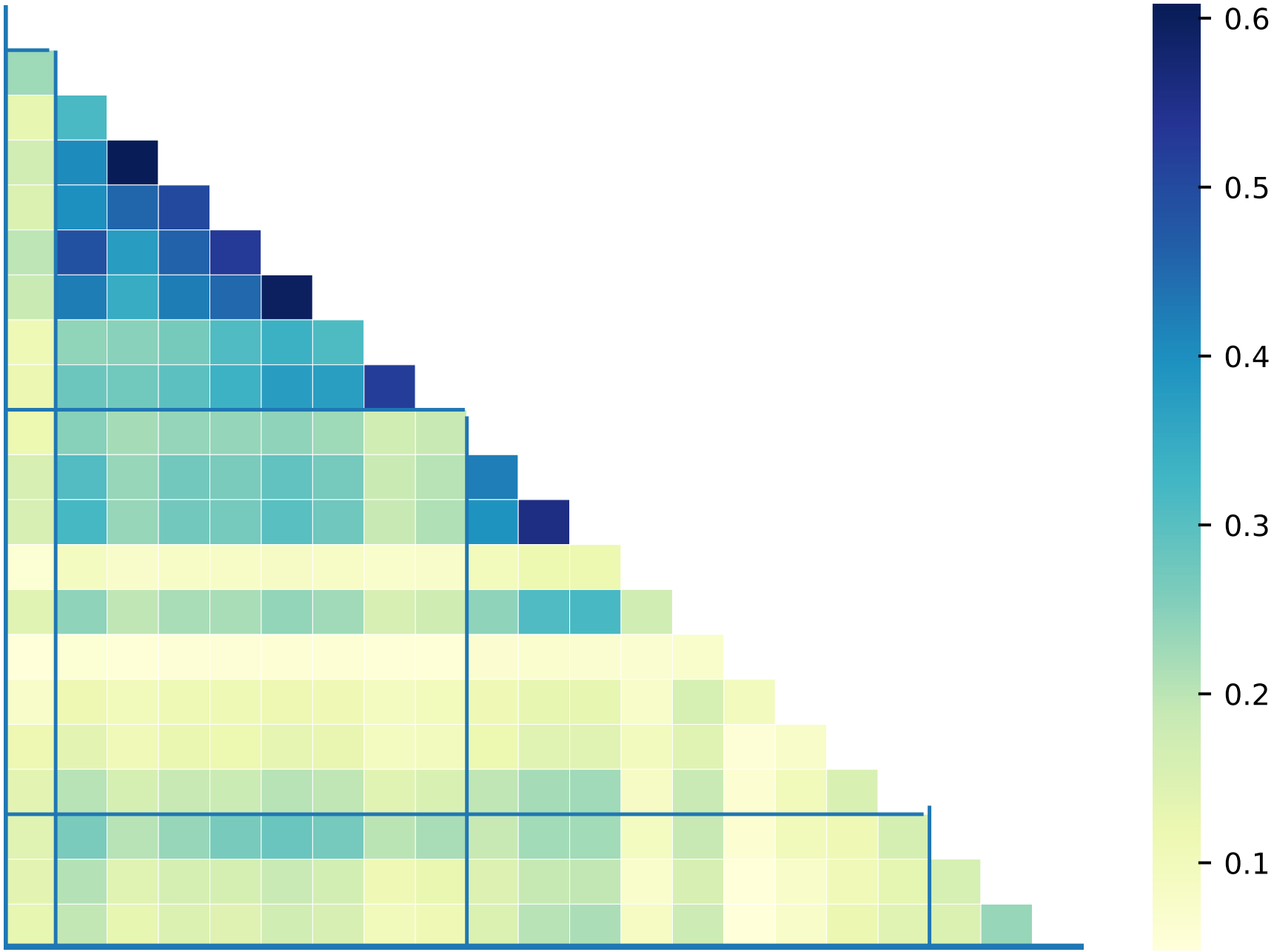
- Embed every sentence in the corpus
- Identify *k* nearest neighbors per embedding model for a fixed set of queries

Nearest neighbor overlap (N2O)



- Compute overlap in nearest neighbors

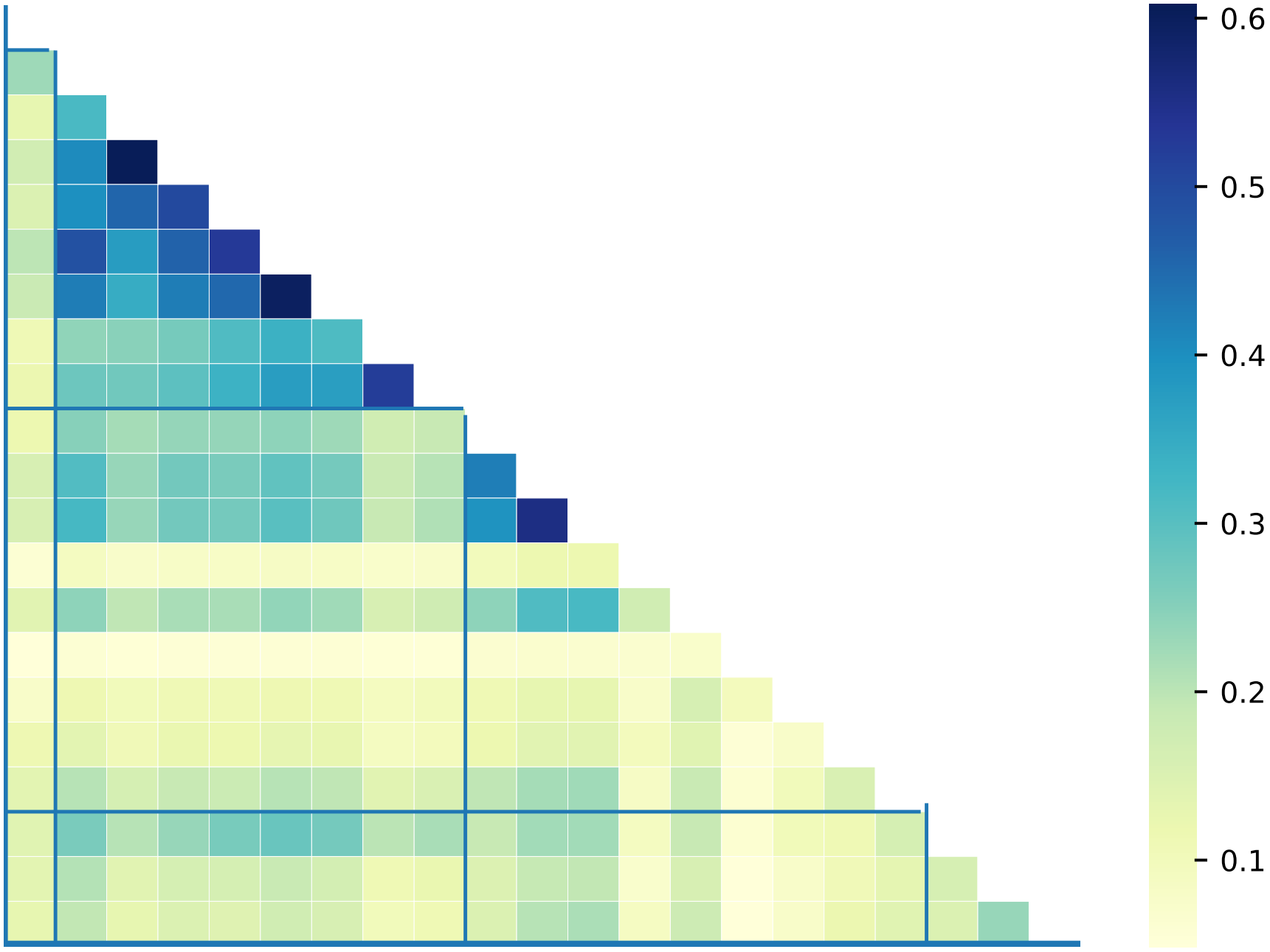
$$\text{N2O} = \frac{2 \text{ common neighbors}}{k = 6} = 0.33$$

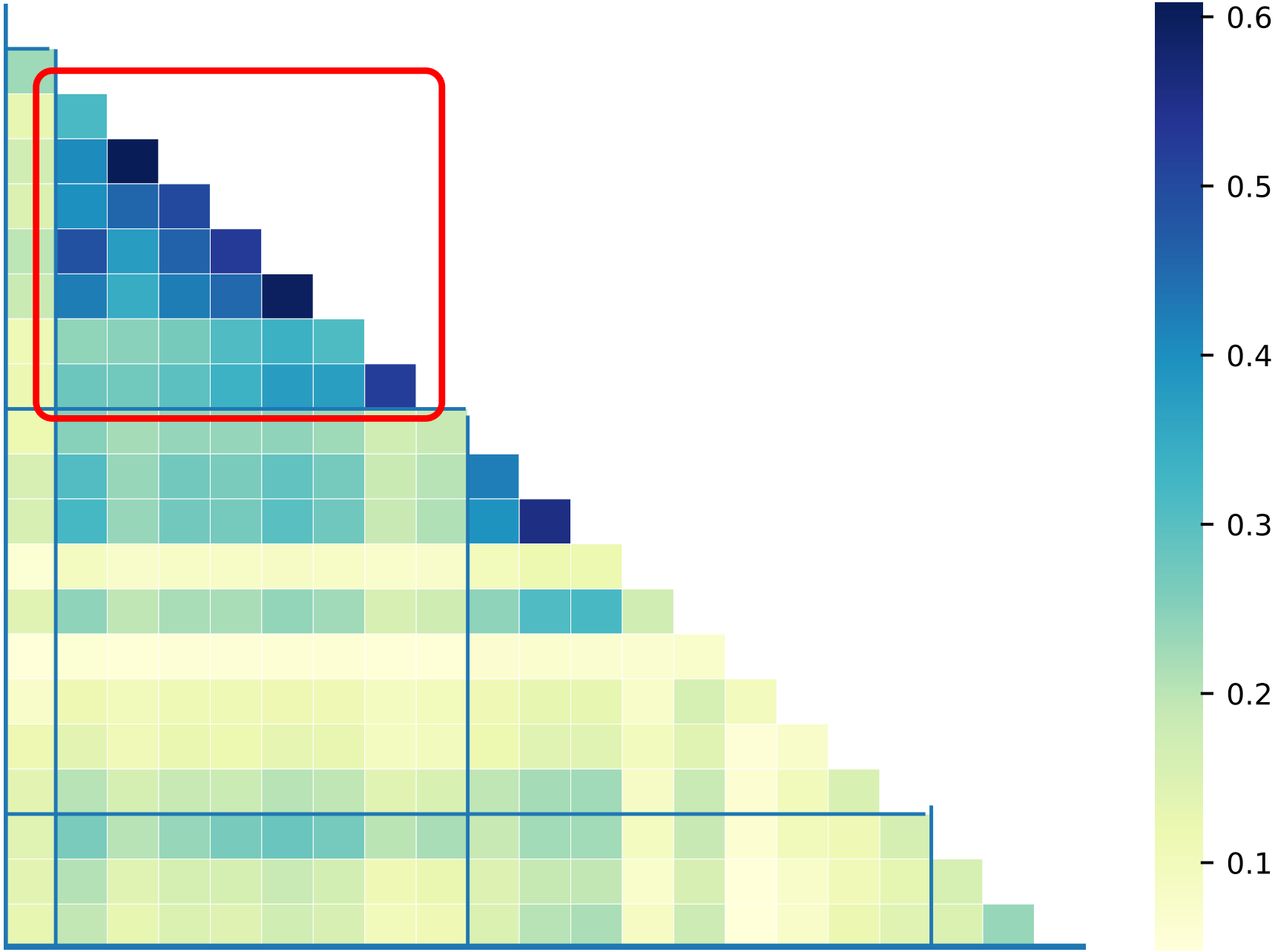

0.6
0.5
0.4
0.3
0.2
0.1

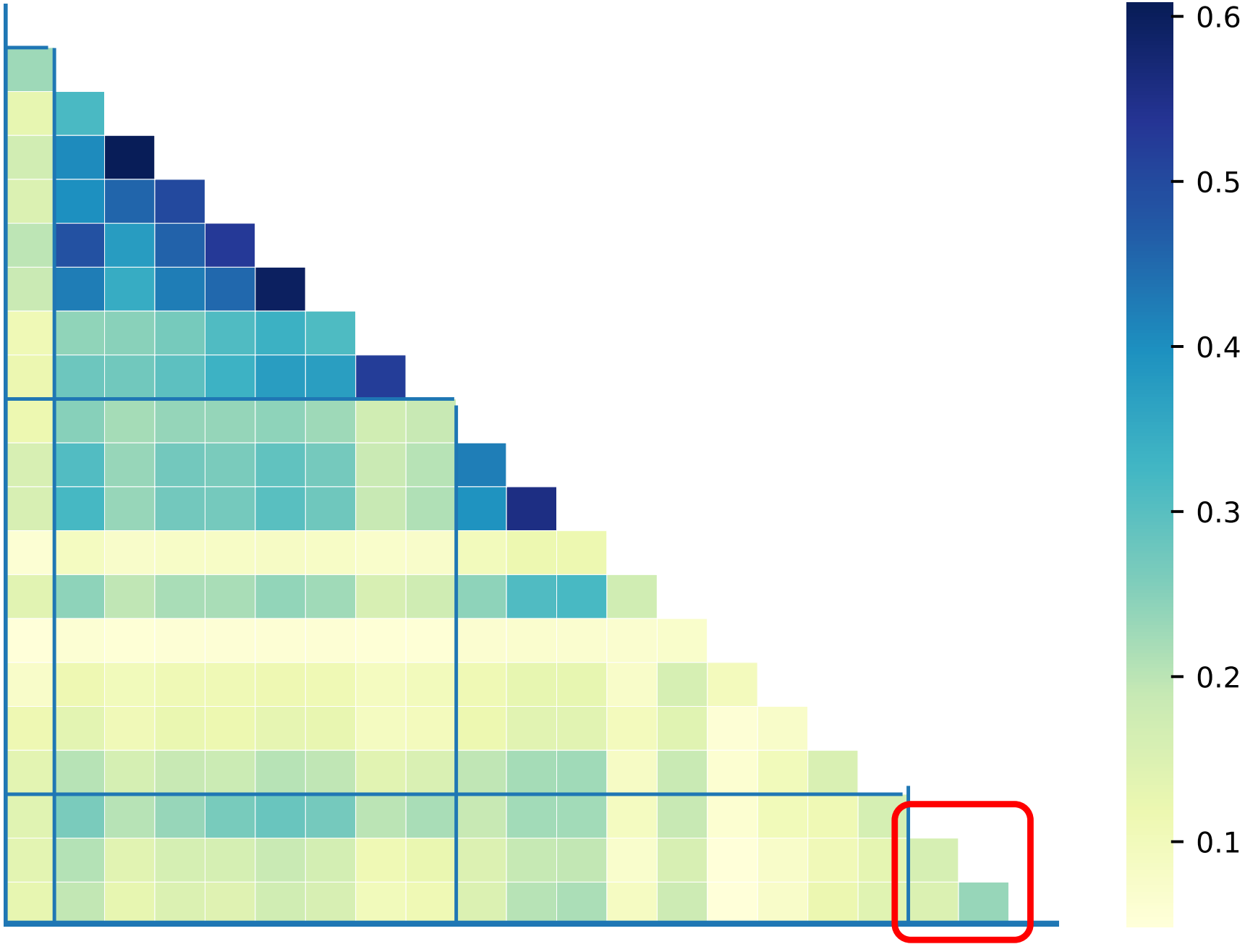
Experimental setup

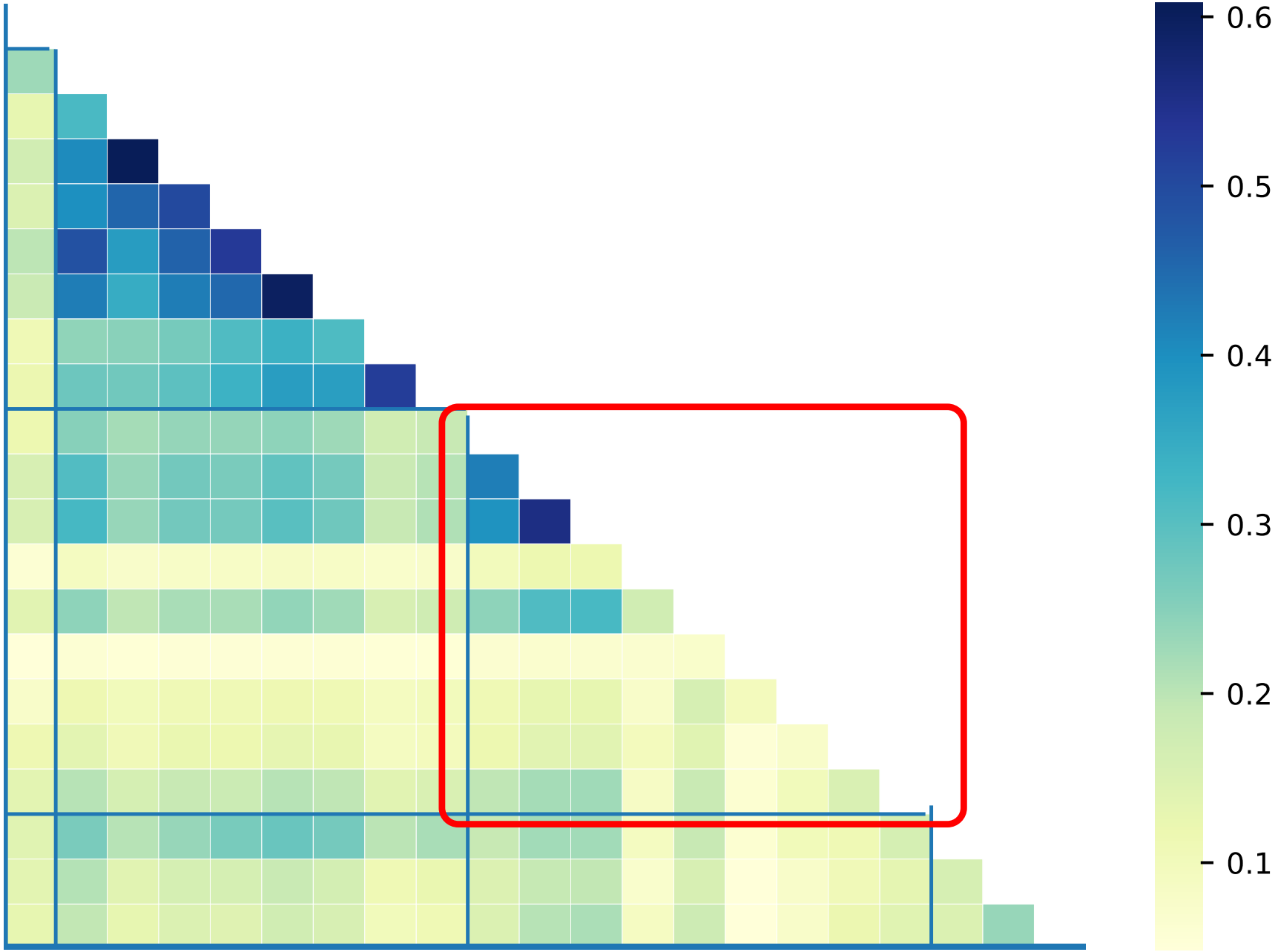
Corpus.	English Gigaword (7 news sources, 2010) approx. 8m unique sentences
Queries.	sample uniformly from news article ledes; $n = 100$
Neighbors.	$k = 50$
Embedding models.	x21

Aside: robust to different query samples & k!


0.6
0.5
0.4
0.3
0.2
0.1


0.6
0.5
0.4
0.3
0.2
0.1


0.6
0.5
0.4
0.3
0.2
0.1


0.6
0.5
0.4
0.3
0.2
0.1

Case study: BERT

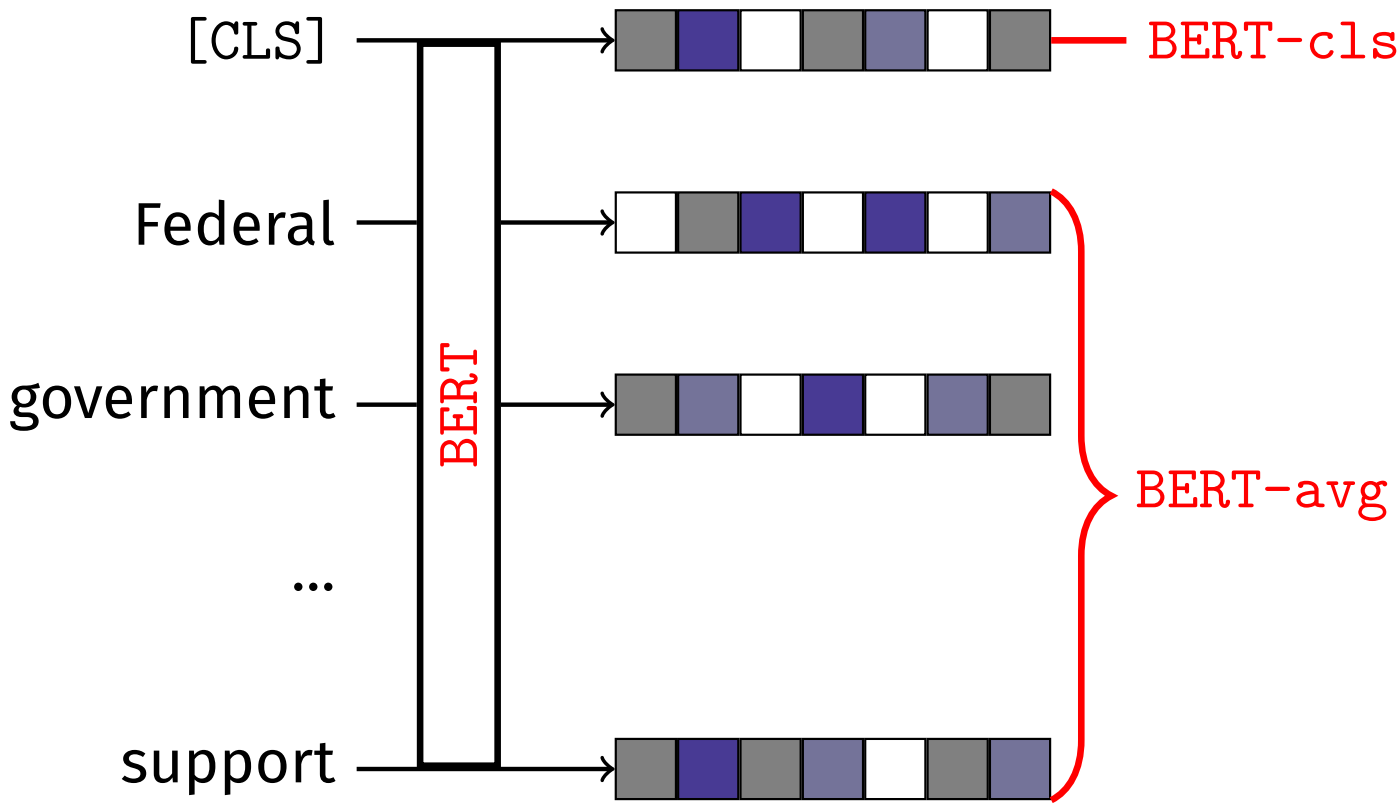
Contextual word embeddings

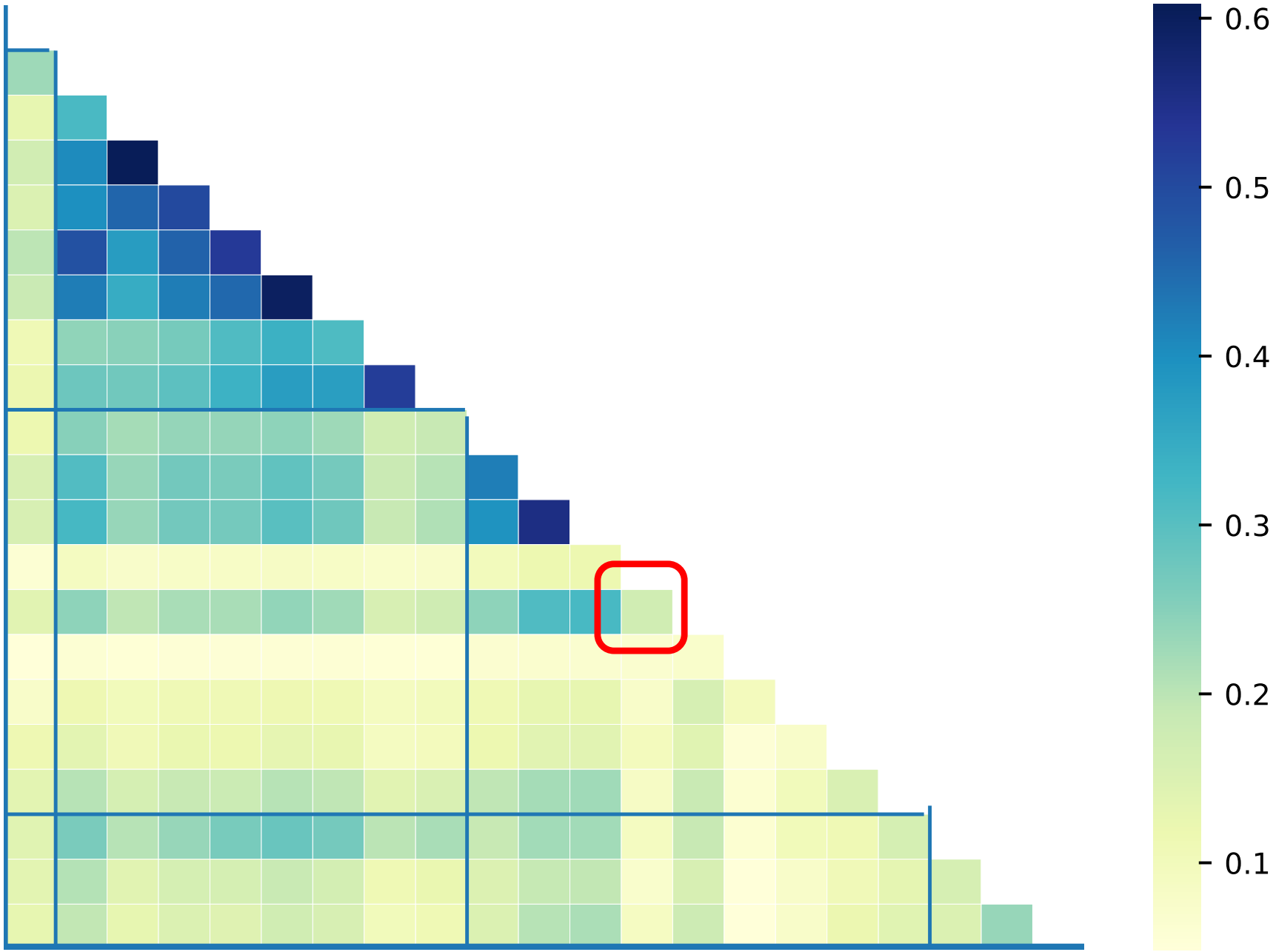
Intuition. Word embeddings should depend not just on the word, but also its context.

Recent popular method: BERT (Devlin et al., 2019)

Contextual word embeddings

How do we get sentence embeddings?




0.6
0.5
0.4
0.3
0.2
0.1

Query.
Britain's biggest mortgage lender says that average house prices fell 3.6 percent in September, but analysts believe the market isn't that weak.

"Popular" neighbor (across all 21 embedding models).
Average house prices in Britain fell 3.6 percent in September from a month earlier, the country's biggest mortgage lender said Thursday, although analysts believe the market isn't that weak.

Query.
Britain's biggest mortgage lender says that average house prices fell 3.6 percent in September, but analysts believe the market isn't that weak.

"Popular" neighbor (across all 21 embedding models).
Average house prices in Britain fell 3.6 percent in September from a month earlier, the country's biggest mortgage lender said Thursday, although analysts believe the market isn't that weak.

Query.

Britain's biggest mortgage lender says that average house prices fell 3.6 percent in September, but analysts believe the market isn't that weak.

Near neighbor for `BERT-cls` *only.*

Japanese consumer prices fell for 13th straight month in March, though the GDP data suggests that deflationary pressures are starting to ease.

Query.

Britain's biggest mortgage lender says that average house prices fell 3.6 percent in September, but analysts believe the market isn't that weak.

Near neighbor for `BERT-cls` *only.*

Japanese consumer prices fell for 13th straight month in March, though the GDP data suggests that deflationary pressures are starting to ease.

Paraphrase $\Leftrightarrow$ nearness?

Exp. use paraphrase pairs from STS (Cer et al. 2018)
"needle in a haystack"

tl;dr. `BERT-cls` does really poorly,
`BERT-avg` does so-so
(aside: there are better models for this)

"Arkansas Supreme Court strikes down execution law."

$\Downarrow$

"Arkansas justices strike down death penalty."

Should I use `BERT-cls` or `BERT-avg` or something else?

It depends on:

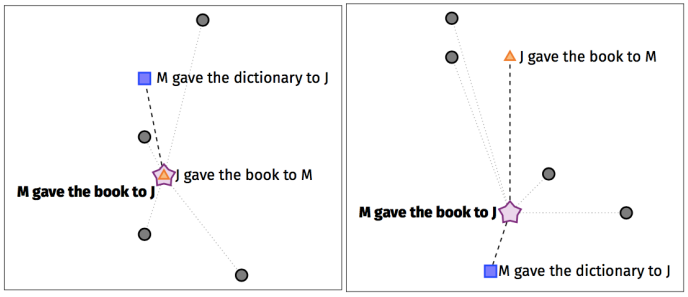
- what kinds of text you're embedding
- how you plan to use the sentence embeddings
- ...and other factors!

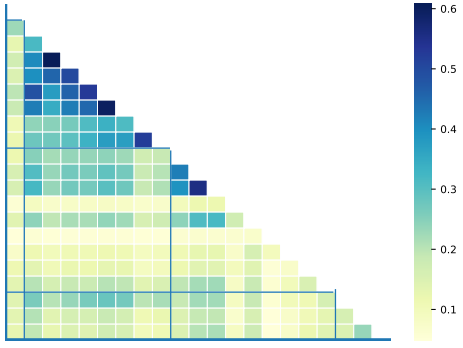
N2O helps guide comparison of embedding models.

N2O: a procedure for comparing sentence embedding models

general experimental results

case study: BERT





Query.
Britain's biggest mortgage lender says that average house prices fell 3.6 percent in September, but analysts believe the market isn't that weak.

Near neighbor for `BERT-cls` *only.*
Japanese consumer prices fell for 13th straight month in March, though the GDP data suggests that deflationary pressures are starting to ease.

Thanks!

Contact: `lucylin@cs.washington.edu`