

Factors Influencing the Surprising Instability of Word Embeddings

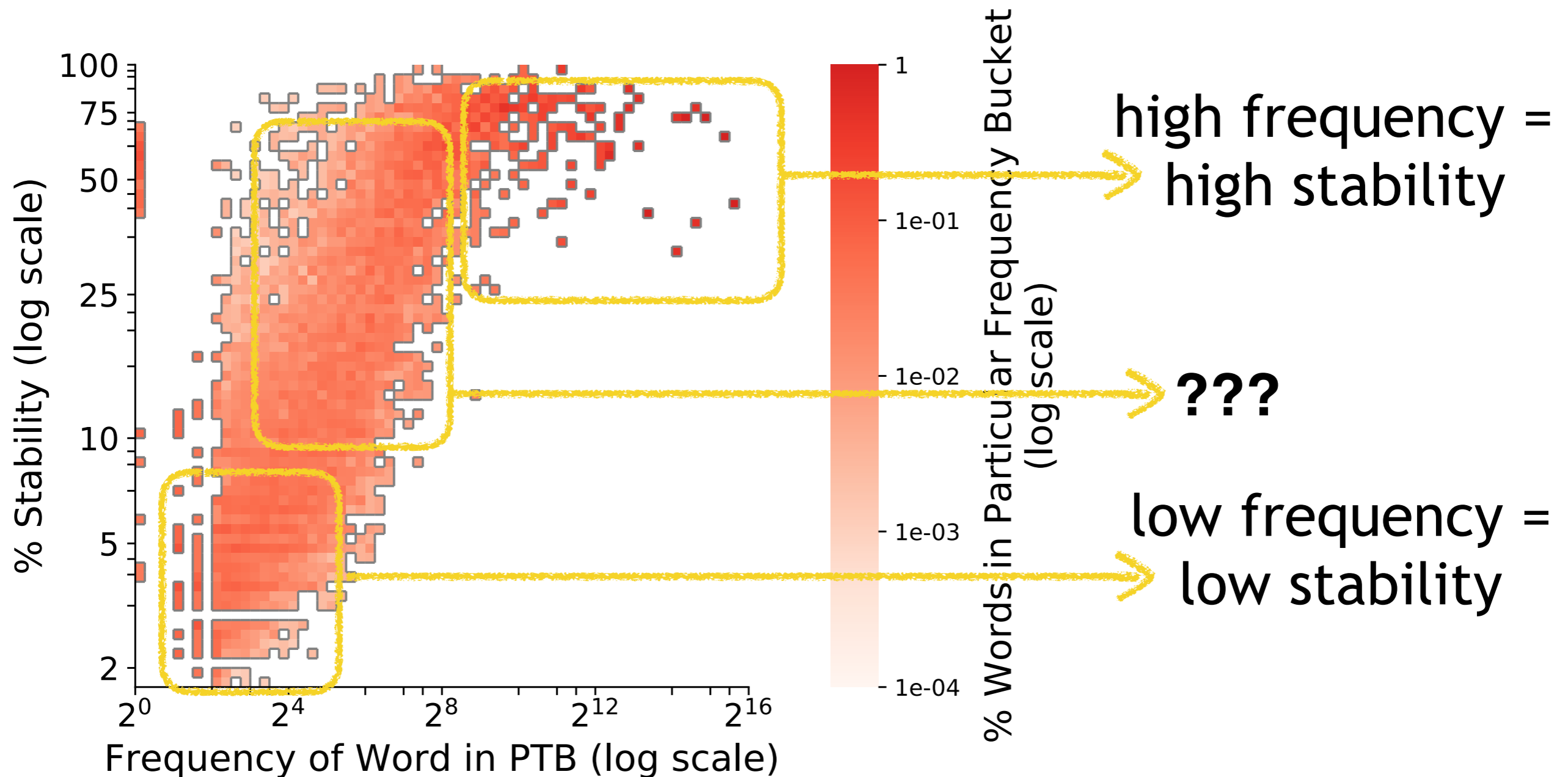
Laura Wendlandt, Jonathan K. Kummerfeld, Rada Mihalcea
University of Michigan



This material is based in part upon work supported by the National Science Foundation (NSF #1344257) and the Michigan Institute for Data Science (MIDAS).

The Problem

Many common embedding algorithms have large amounts of instability.



The Model

A ridge regression model that aims to predict the stability of a word given: (1) word properties; (2) data properties; and (3) algorithm properties

Data:

- New York Times (NYT)— six domains: U.S., New York and Region, Business, Arts, Sports, all NYT
- Europarl

Algorithms:

- *word2vec* skip-gram model
- GloVe
- PPMI

1. Frequency is not a major factor in stability.

Model with frequency:

R² score of 0.301

Model without frequency:

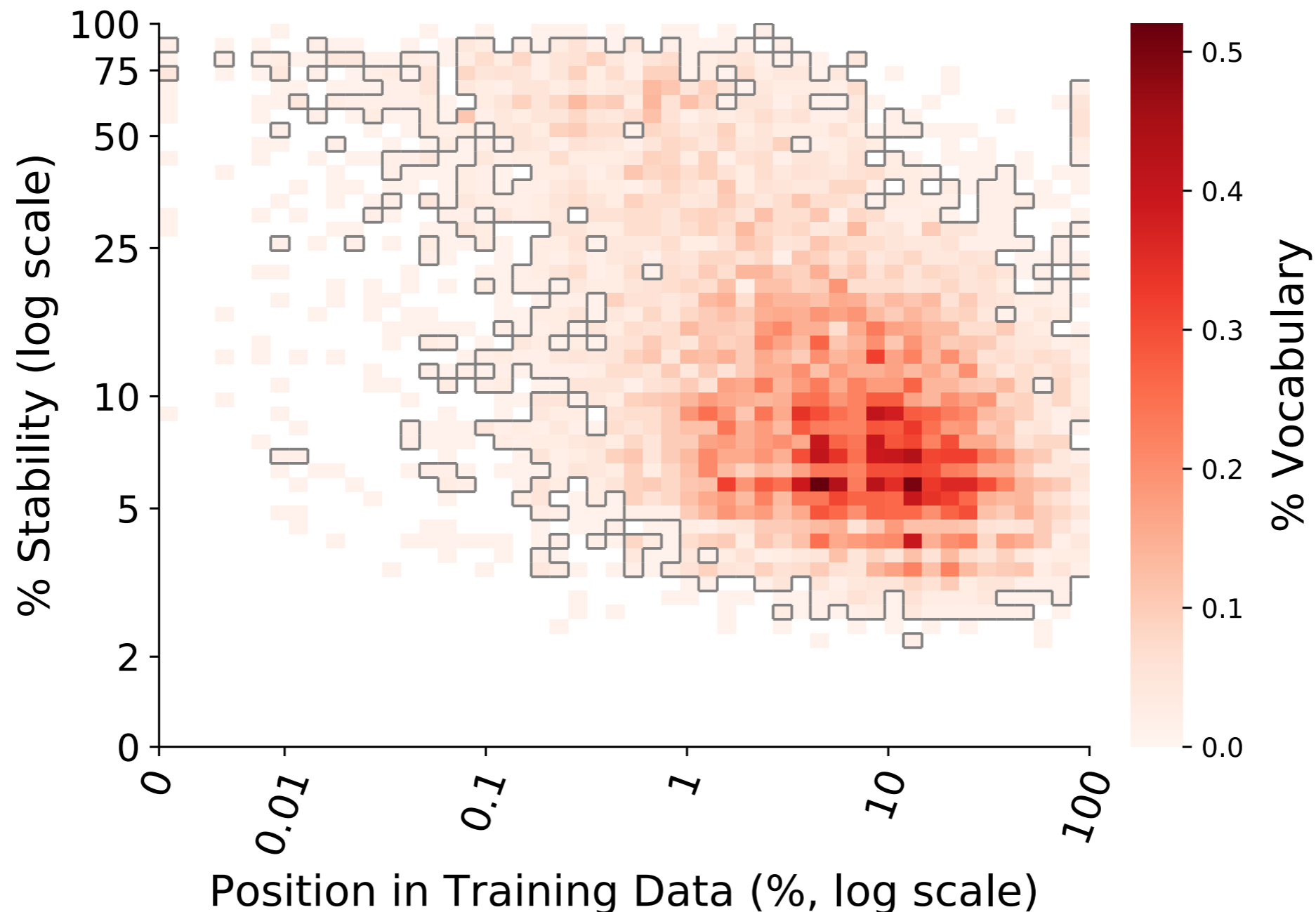
R² score of 0.301

Model with **only** frequency:

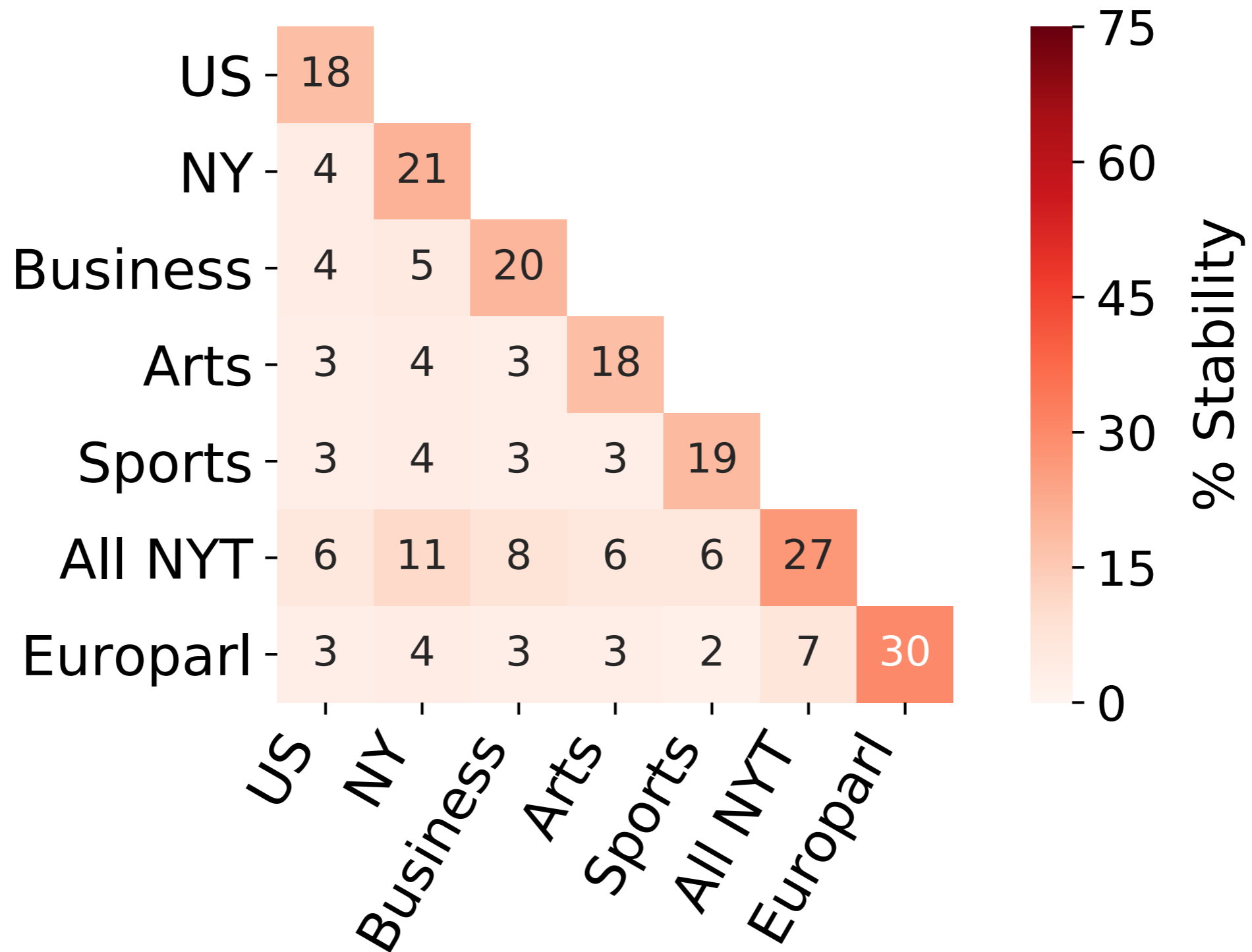
R² score of 0.008

2. Curriculum learning is important.

Curriculum learning = order of training data given to an algorithm



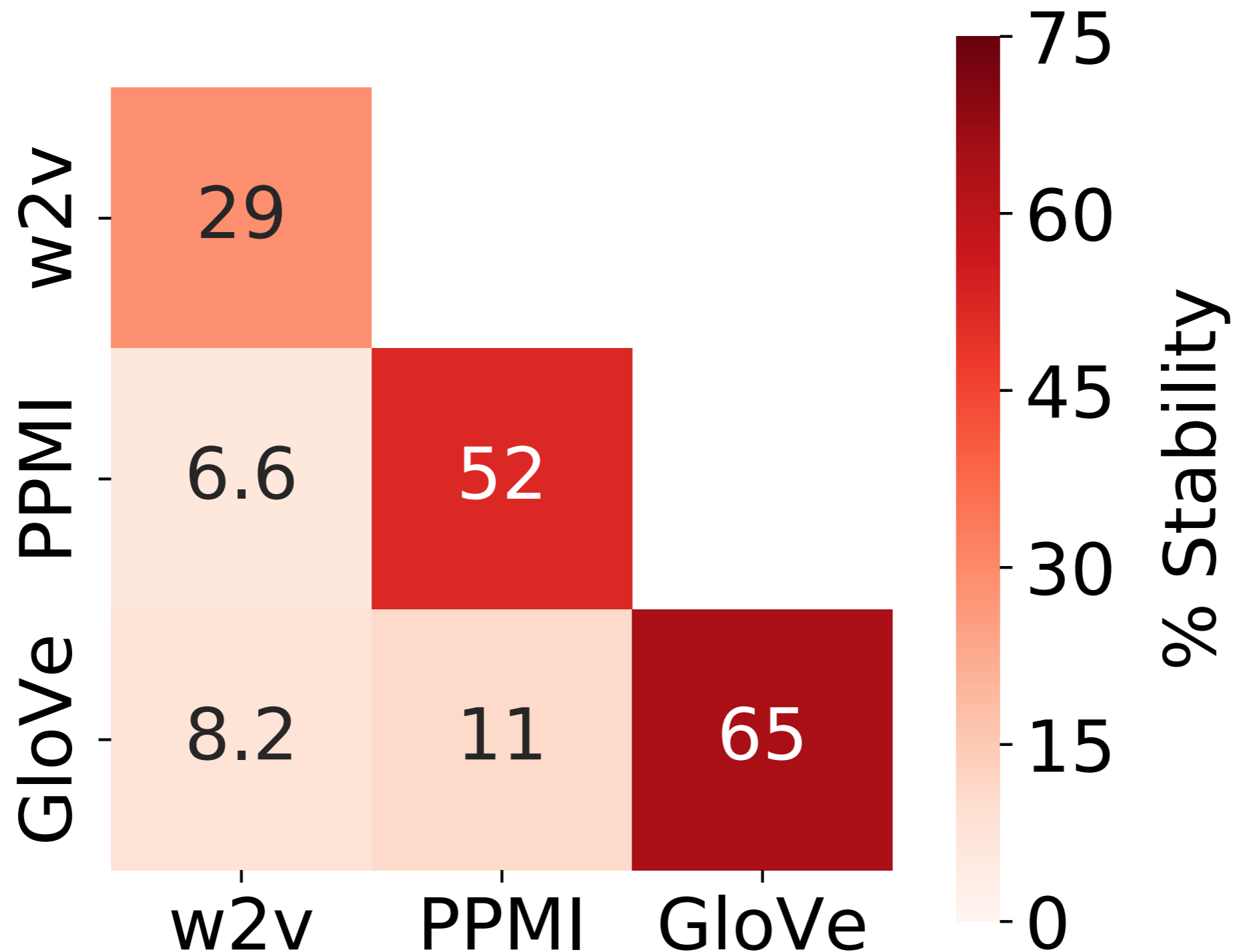
3. Stability within domains is greater than across domains.



4. POS is one of the biggest factors in stability.

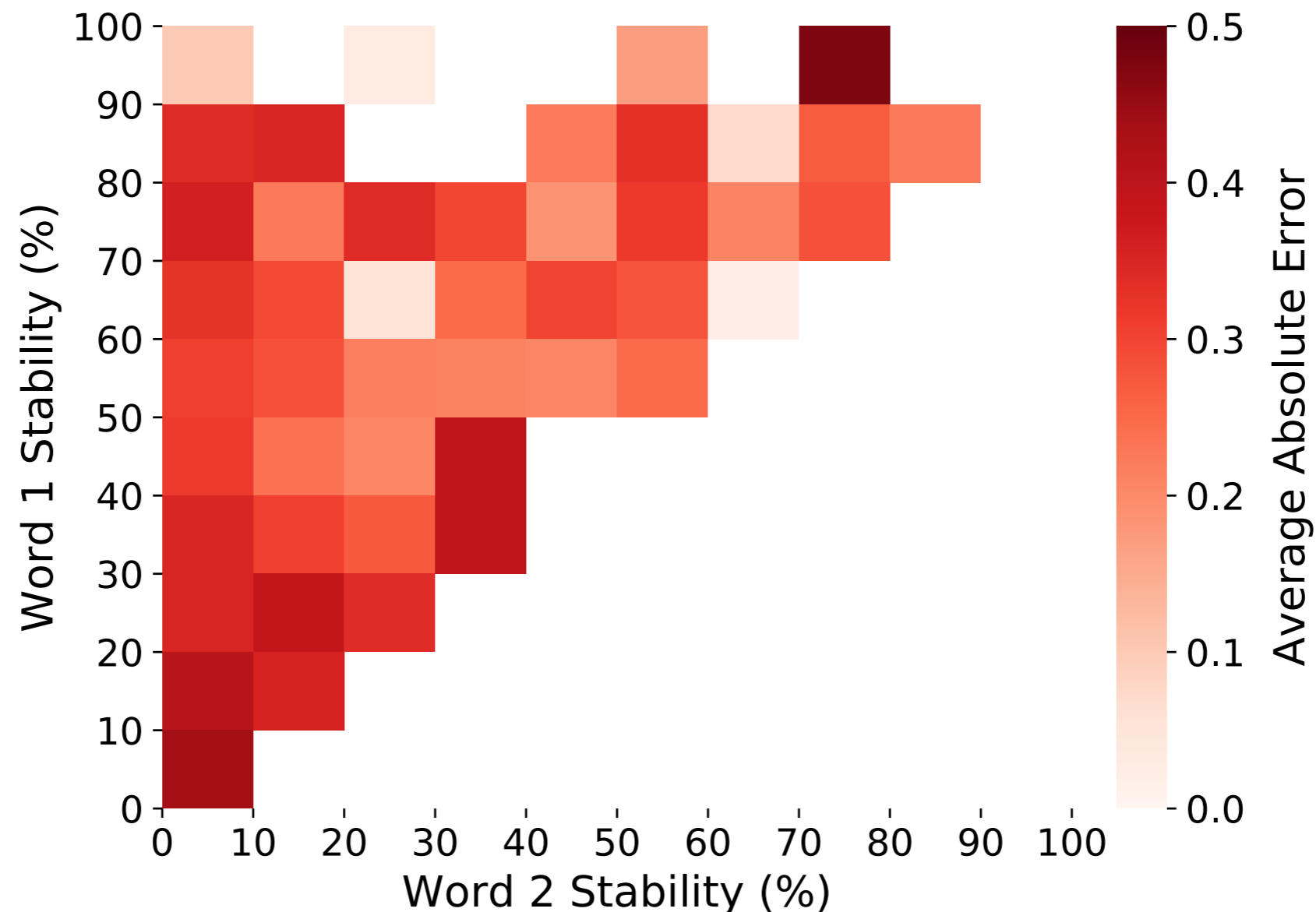
<i>POS</i>	<i>Avg. Stability</i>
Numeral	47%
Verb	31%
Determiner	31%
Adjective	31%
Noun	30%
Adverb	29%
Pronoun	29%
Conjunction	28%
Particle	26%
Adposition	25%

5. Overall, GloVe is the most stable embedding algorithm.



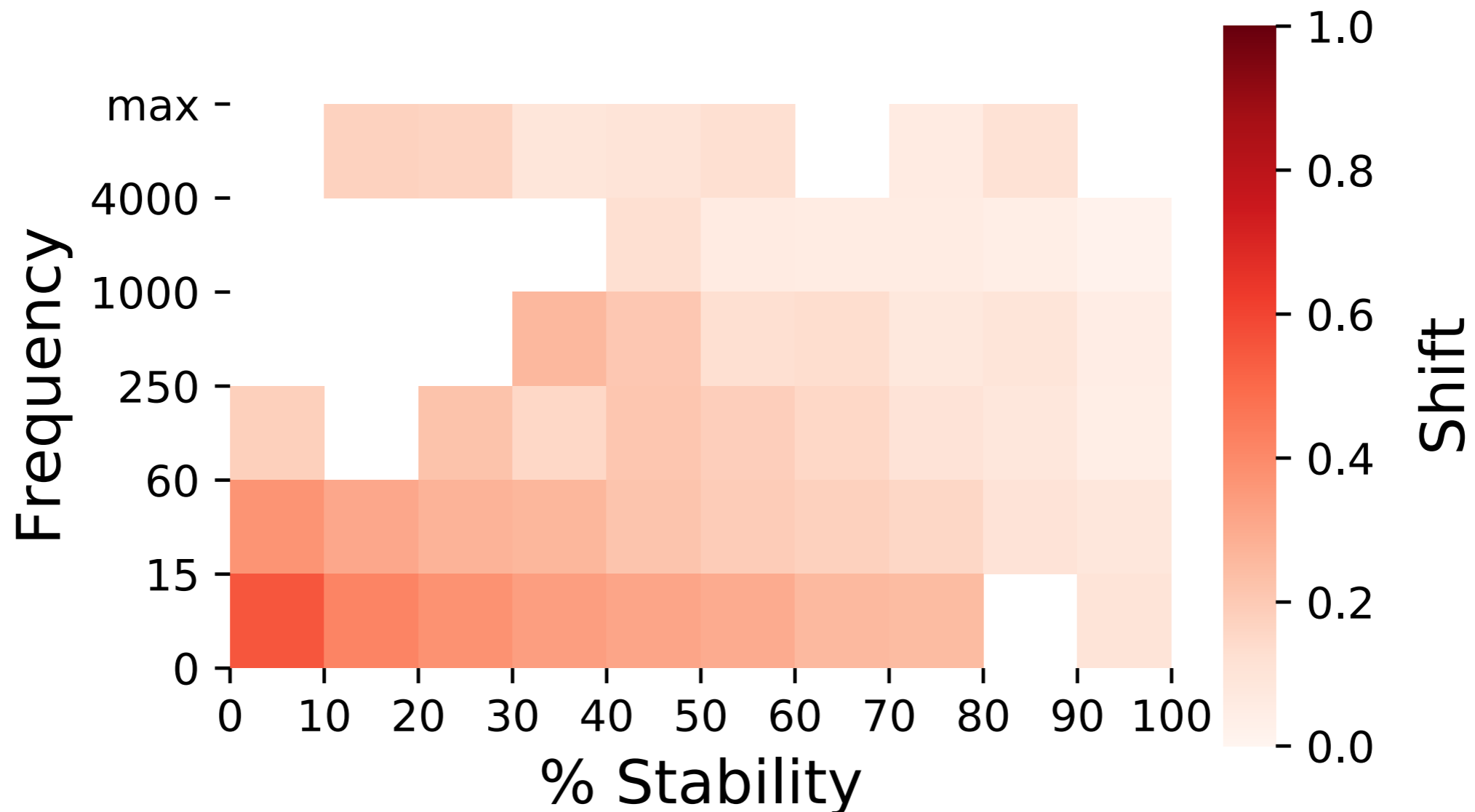
6. Stability affects some downstream tasks.

Word stability correlates slightly with performance on word similarity tasks.



6. Stability affects some downstream tasks.

For POS tagging using an LSTM, the LSTM compensates for instability by shifting unstable word vectors.



Final Thoughts

- *Use GloVe*
- *Learn a good curriculum for word2vec*
- *Use in-domain embeddings whenever possible*

Download our code:

`http://lit.eecs.umich.edu/downloads.html`