On the Downstream Performance of Compressed Word Embeddings

Avner May, Jian Zhang, Tri Dao, Chris Ré
Stanford University
Word Embeddings
Word Embeddings

Important for strong NLP performance
Word Embeddings

- Important for strong NLP performance
- Take a lot of memory
Word Embedding Compression
What determines whether a compressed embedding matrix will perform well on downstream tasks?
What determines whether a compressed embedding matrix will perform well on downstream tasks?
What determines whether a compressed embedding matrix will perform well on downstream tasks?
What determines whether a compressed embedding matrix will perform well on downstream tasks?
Motivating Observation

Existing ways of measuring compression quality often fail to explain relative downstream performance.
Motivating Observation

*Existing ways* of measuring compression quality often *fail to explain* relative downstream performance.
Motivating Observation

Existing ways of measuring compression quality often fail to explain relative downstream performance.
Our Contributions: Outline
Our Contributions: Outline

1. Define a new measure of compression quality.
Our Contributions: Outline

1. Define a **new measure** of compression quality.

2. Prove **generalization bounds** using this measure.
Our Contributions: Outline

1. Define a **new measure** of compression quality.

2. Prove **generalization bounds** using this measure.

3. Show strong **empirical correlation** w. downstream performance.
Our Contributions: Outline

① Define a **new measure** of compression quality.

② Prove **generalization bounds** using this measure.

③ Show strong **empirical correlation** w. downstream performance.

④ Use measure to **select** compressed embeddings.
Our Contributions: Outline

1. Define a new measure of compression quality.
2. Prove generalization bounds using this measure.
3. Show strong empirical correlation w. downstream performance.
4. Use measure to select compressed embeddings.

Up to 2x lower selection error rates than the next best measure.
Defining the Measure: Intuition from Linear Regression

**Observation:** Predictions are determined by data matrix’s *left singular vectors.*
Defining the Measure: Intuition from Linear Regression

Observation:
Predictions are determined by data matrix’s left singular vectors.
Defining the Measure: Intuition from Linear Regression

**Observation:**
Predictions are determined by data matrix’s *left singular vectors.*
Defining the Measure: Intuition from Linear Regression

Observation:
Predictions are determined by data matrix’s left singular vectors.
Defining the Measure: Intuition from Linear Regression

**Observation:** Predictions are determined by data matrix’s *left singular vectors*.

![Diagram](image)

- Embed. matrix
- Singular Value Decomposition
- Regression label $Y$
- Project $y$ onto span of *left singular vectors*
Defining the Measure: Eigenspace Overlap Score (EOS)

**Intuition:**
Measures similarity between the span of *left singular vectors.*
Defining the Measure: Eigenspace Overlap Score (EOS)

**Intuition:**
Measures similarity between the span of *left singular vectors.*

$$\text{EOS}(d, \cdot) = \frac{1}{d} \left\| F \right\|_2^2$$
Theoretical Results: Linear Regression

Theorem (informal):
Expected difference in test mean-squared error attained by compressed vs. uncompressed embeddings is determined by EOS.
Theoretical Results: Linear Regression

Theorem (informal):
Expected difference in *test mean-squared error* attained by *compressed* vs. *uncompressed* embeddings is *determined by EOS*.

Higher EOS
Theoretical Results: Linear Regression

Theorem (informal): Expected difference in test mean-squared error attained by compressed vs. uncompressed embeddings is determined by EOS.

Higher EOS \rightarrow Better downstream performance
Empirical Correlation: Beyond Linear Regression

EOS attains strong correlation with downstream model accuracy.
Empirical Correlation: Beyond Linear Regression

EOS attains strong correlation with downstream model accuracy.
Empirical Correlation: Beyond Linear Regression

EOS attains strong correlation with downstream model accuracy.
Empirical Correlation: Beyond Linear Regression

EOS attains **strong correlation** with downstream **model accuracy**.

Higher accuracy

Higher quality

---

EOS as a Selection Criterion

EOS attains \textit{up to 2x lower selection} error rates than 2\textsuperscript{nd} best.
EOS as a Selection Criterion

EOS attains **up to 2x lower selection** error rates than 2\textsuperscript{nd} best.

![Graph showing selection error rates for different NLP tasks]

EOS as a Selection Criterion

EOS attains \textit{up to 2x lower selection} error rates than 2\textsuperscript{nd} best.

Our Contributions: Summary
Our Contributions: Summary

① Defined a new measure of compression quality.
Our Contributions: Summary

1. Defined a **new measure** of compression quality.

2. Proved **generalization bounds** using this measure.
Our Contributions: Summary

1. Defined a **new measure** of compression quality.

2. Proved **generalization bounds** using this measure.

3. Showed strong **empirical correlation** w. downstream perf.
Our Contributions: Summary

1. Defined a new measure of compression quality.

2. Proved generalization bounds using this measure.

3. Showed strong empirical correlation w. downstream perf.

4. Used measure to select compressed embeddings.
THANK YOU!

Poster #185, 5-7 pm today!

Code: https://github.com/HazyResearch/smallfry
E-mail: avnermay@cs.stanford.edu