Understanding the Downstream Instability of Word Embeddings

Megan Leszczynski, Avner May, Jian Zhang, Sen Wu, Chris Aberger, Chris Ré

Stanford University
Motivation

Recommend new content

Detect the latest spam

Learn new words

Why retrain?

Changing distribution of popular videos

New spam techniques

Out-of-vocabulary words

Model freshness is necessary for user satisfaction in many products.
Google retrains their app store Google Play models every day, and Facebook retrained search models every hour.

But model training can be unstable...

Prediction churn

Challenges of Instability

Debugging

Consistent user-experience

Model dependencies

Research reliability
Problem Setting: Embedding Server

How does the embedding instability propagate to these tasks?
Key takeaway:
Stability–memory tension

With the right understanding, we can improve stability by over 30% – in the same amount of memory.
Outline

Q: How do we define downstream instability?
   A: % prediction disagreement

Q: What embedding hyperparameters impact downstream instability?
   A: hyperparameters related to memory

Q: How can we theoretically understand downstream instability?
   A: using our eigenspace instability measure (EIS)

Q: How can we select embedding hyperparameters to minimize instability?
   A: using the EIS (or k-NN) measures
Definition: Downstream Instability

Downstream instability = % prediction disagreement between models trained on a pair of embeddings

Metrics like instability are important for modularity.
Outline

Q: How do we define downstream instability?
   A: % prediction disagreement

Q: What embedding hyperparameters impact downstream instability?
   A: hyperparameters related to memory

Q: How can we theoretically understand downstream instability?
   A: using our eigenspace instability measure (EIS)

Q: How can we select embedding hyperparameters to minimize instability?
   A: using the EIS (or k-NN) measures
Hyperparameters that Impact Memory

\[ \text{Dimension} \quad \frac{\# \text{ features / word}}{\# \text{ bits / feature}} = \text{Memory} \]

<table>
<thead>
<tr>
<th>Uniform Quantization</th>
<th>32-bit</th>
<th>1-bit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interval: [-0.1, 0.1]</td>
<td>-0.03</td>
<td>-0.1</td>
</tr>
<tr>
<td>[0.04, 0.1]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Impact of Dimension

Sentiment Analysis

NER

Dimension

Downstream Instability
Impact of Precision

Sentiment Analysis

NER

Precision
Downstream Instability
Stability-Memory Tradeoff

Sentiment Analysis

NER

SST-2

CoNLL-2003

Memory (Bits / Word)

% Disagreement

Memory (Bits / Word)

% Disagreement

11%

↑ Memory  ↓ Downstream Instability
Q: How do we define downstream instability?
   A: % prediction disagreement

Q: What embedding hyperparameters impact downstream instability?
   A: hyperparameters related to memory

Q: How can we theoretically understand downstream instability?
   A: using our eigenspace instability measure (EIS)

Q: How can we select embedding hyperparameters to minimize instability?
   A: using the EIS (or k-NN) measures
Goal: Embedding distance measure

The measure must relate the distance between the embeddings to the downstream instability.
Eigenspace Instability Measure (EIS)

Key insight:

The predictions of a linear regression model trained on an embedding $X$ depend on the \textbf{left singular vectors} of $X$.\footnote{May et al. On the downstream performance of compressed word embeddings. NeurIPS, 2019.}

\[
\text{Emb}(X) = S U V^T
\]
Eigenspace Instability Measure (EIS)

- EIS measures the similarity of the left singular vectors of two embeddings

For embeddings $X$ and $\tilde{X}$,

$$\text{EIS} (X, \tilde{X}) = \text{similarity}(U, \tilde{U})$$

- Can be computed in time $O(nd^2)$
  - $n$ is the size of vocabulary and $d$ is the dimension
Eigenspace Instability Measure (EIS)

Theorem (informal):

EIS is equal to the expected mean-squared difference between the predictions of the linear models trained on $X$ and $\tilde{X}$.

Direct theoretical connection between the EIS measure and the downstream instability.
Outline

Q: How do we define downstream instability?
   A: % prediction disagreement

Q: What embedding hyperparameters impact downstream instability?
   A: hyperparameters related to memory

Q: How can we theoretically understand downstream instability?
   A: using our eigenspace instability measure (EIS)

Q: How can we select embedding hyperparameters to minimize instability?
   A: using the EIS (or k-NN) measures
Embedding measure for downstream instability?

- EIS measure
- k-NN measure [1,2,3]
- Semantic displacement (SD) [4]
- PIP loss [5]
- Eigenspace overlap (EO) [6]

EIS and k-NN measures strongly correlate with downstream instability.
Selection Task Setup

- Use embedding distance measure to select hyperparameters for a fixed memory budget
- Record the difference in downstream instability to the oracle hyperparameters
Selection Task Results

EIS and k-NN measures outperform other measures as selection criteria.
Our theoretically grounded measure improves the stability **up to 34%** over a full precision baseline in the same amount of memory.
Stability-Memory Tension on KG Embeddings

Memory (Bits / Vector)

Unstable-Rank@10 (%)

100
80
60
40
20
10^1 10^2 10^3 10^4

Link Prediction

b=1
b=2
b=4
b=8
b=16
b=32

% Disagreement

35
30
25
20
15
10^1 10^2 10^3 10^4

Triplet Classification

b=1
b=2
b=4
b=8
b=16
b=32

Memory

Downstream Instability
Conclusion

- Exposed a **stability-memory tradeoff** for word embeddings.
- Proposed the **EIS measure to understand** downstream instability.
- Evaluated measures for **hyperparameter selection** to minimize instability.

*Check out the paper for extended experiments with more embedding algorithms and downstream tasks!*

**Code:**

**Comments or Questions:**
mleszczy@stanford.edu