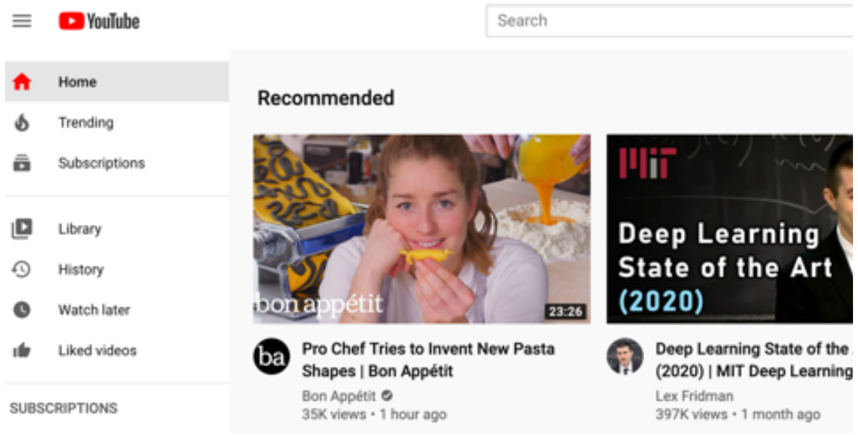
# Understanding the Downstream Instability of Word Embeddings

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

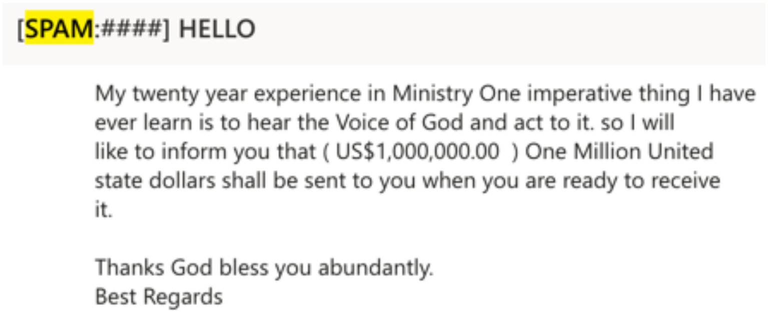
Stanford University

# Motivation

Recommend new content




Detect the latest spam




Learn new words


***Why retrain?***

*Changing distribution of popular videos*

*New spam techniques*

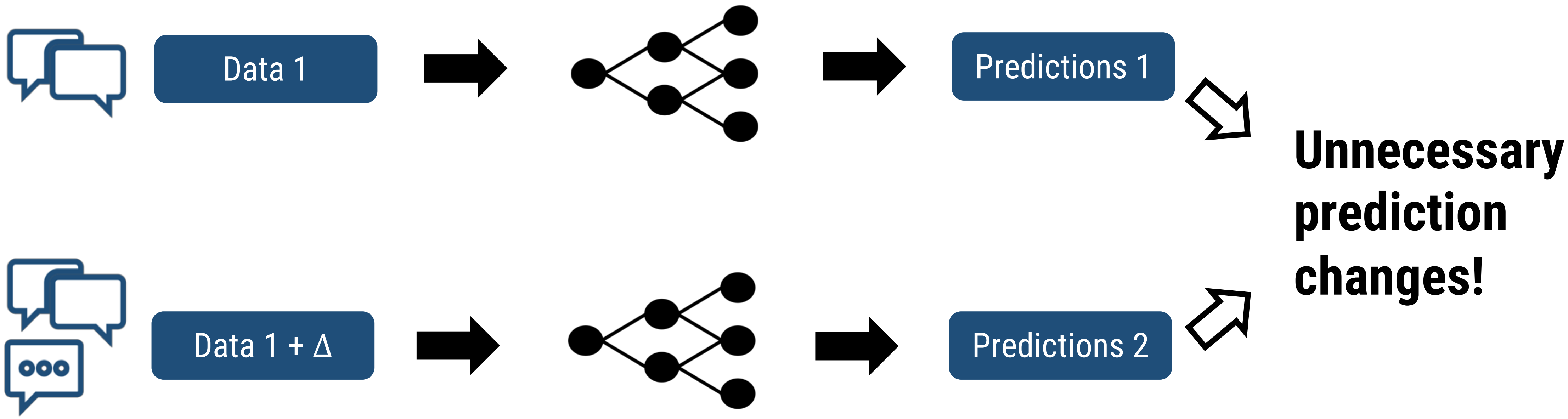
*Out-of-vocabulary words*

**Model freshness is necessary for user satisfaction in many products.**

Google retrains their app store Google Play models *every day*, and Facebook retrains search models *every hour*.

[1] Baylor et al. TFX: A TensorFlow-Based Production-Scale Machine Learning Platform. KDD, 2017.
[2] Hazelwood et al. Applied Machine Learning at Facebook: A Datacenter Infrastructure Perspective. HPCA, 2018.

# But model training can be unstable…




***Prediction churn***

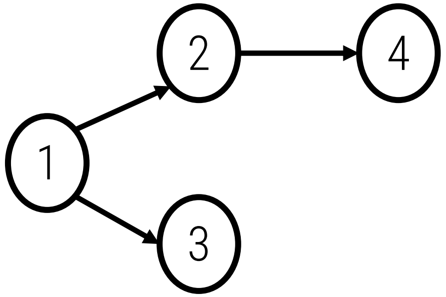
[1] Cormier et al. Launch and Iterate: Reducing Prediction Churn. NeurIPS, 2016.

# Challenges of Instability

Debugging

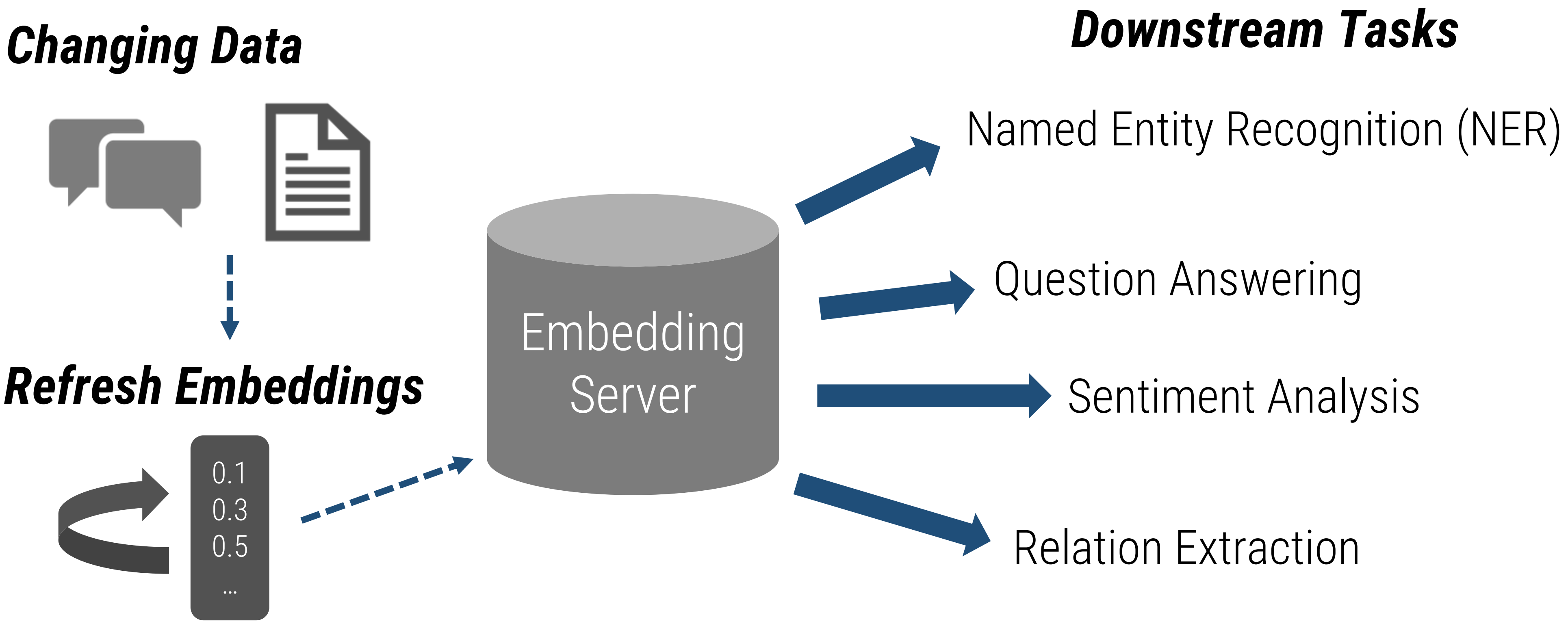
Consistent user-experience

Model dependencies




Research reliability

# Problem Setting: Embedding Server




**Embeddings are shared among downstream tasks.**

**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

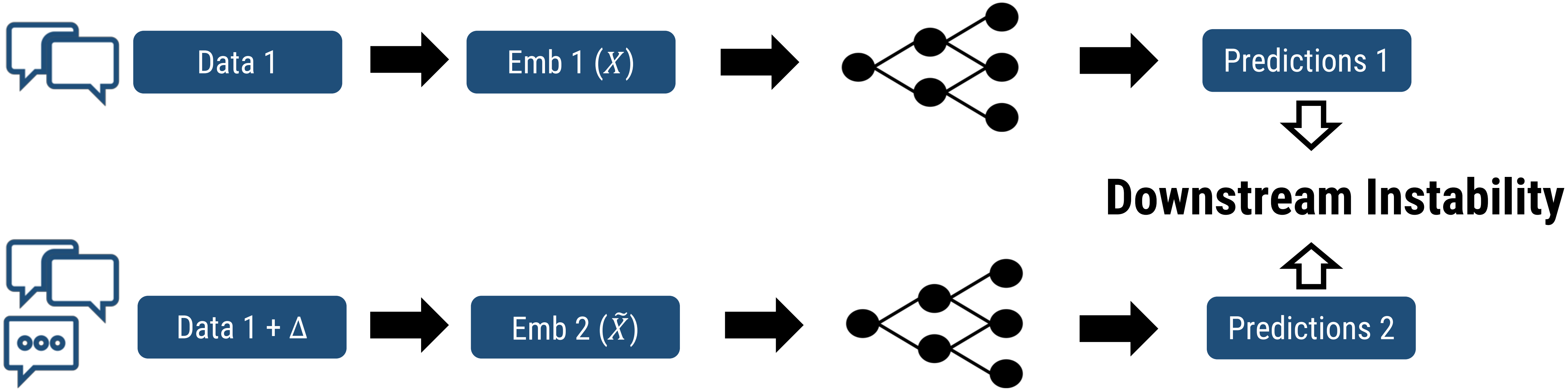
Q: How can we *theoretically understand* downstream instability?

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

↑ **Dimension** ↑ **Precision** = ↑ **Memory**

\# features / word      # bits / feature

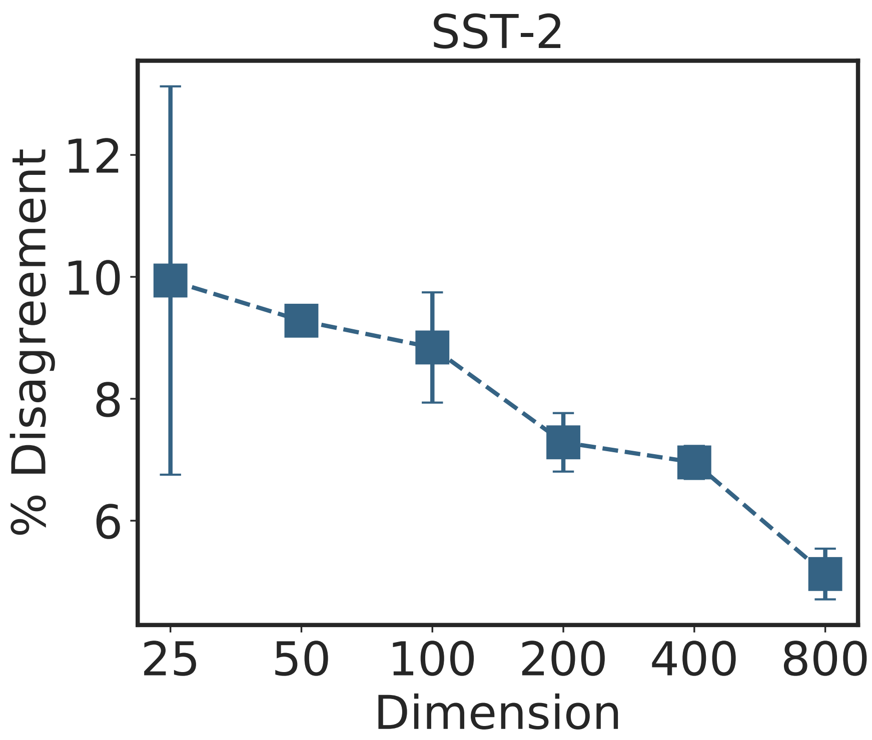
| *Uniform Quantization* | 32-bit | | 1-bit |
|---|---|---|---|
| | **0.04** | | **0.1** |
| Interval: | **-0.03** | → | **-0.1** |
| [-0.1, 0.1] | **-0.08** | | **-0.1** |

↓ **Downstream Instability**

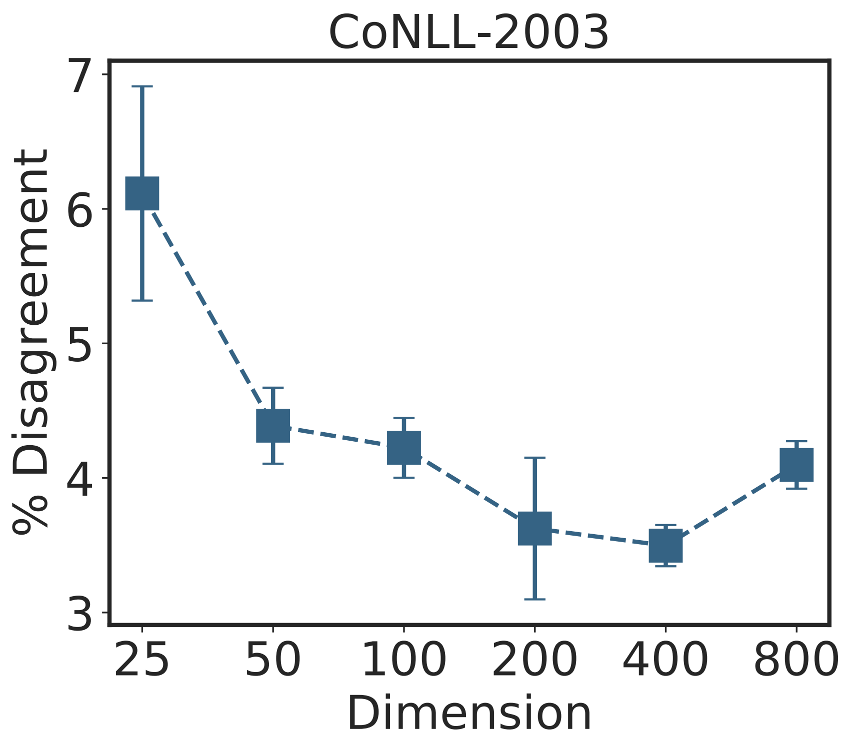
[1] May et al. On the downstream performance of compressed word embeddings. NeurIPS, 2019.

# Impact of Dimension

Sentiment Analysis

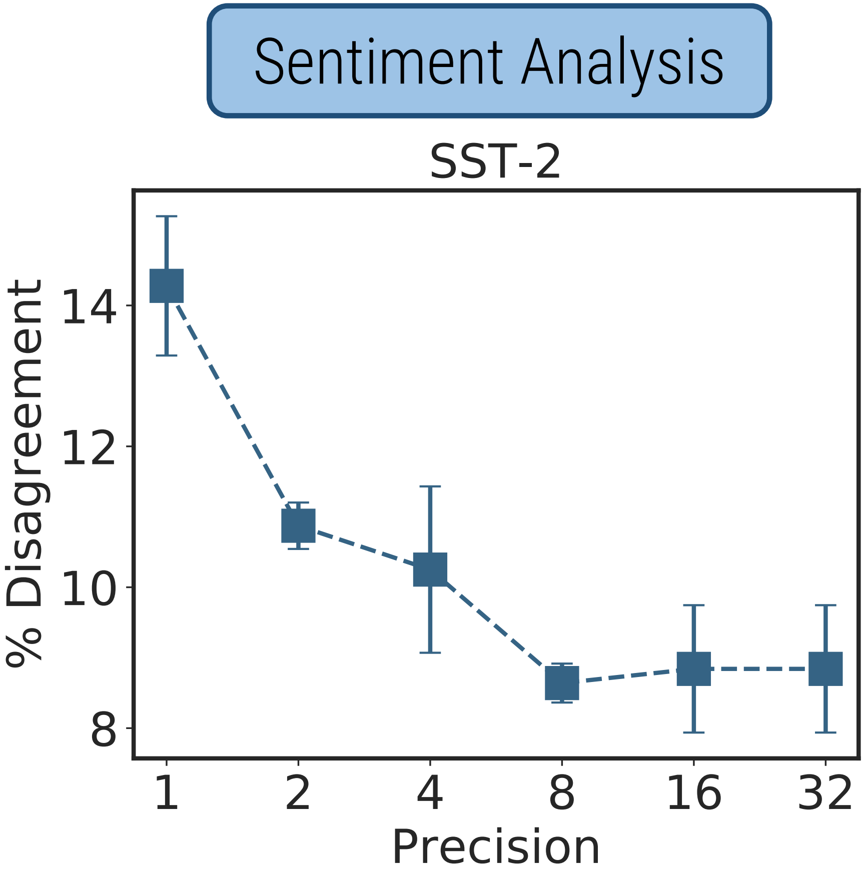


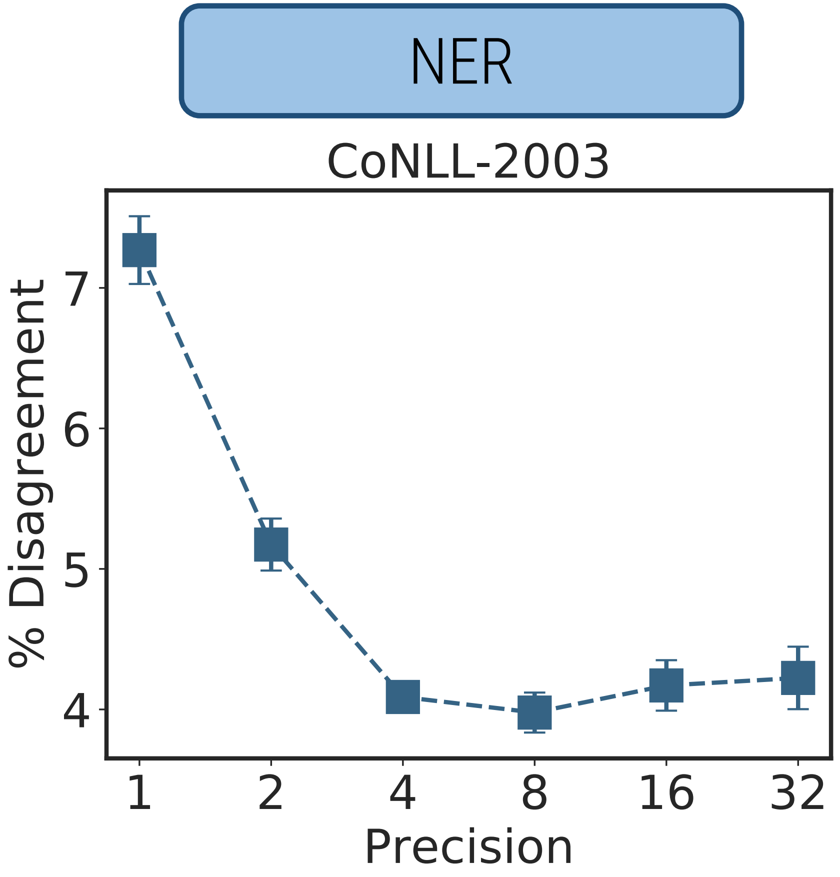

NER




**Dimension** **Downstream Instability**

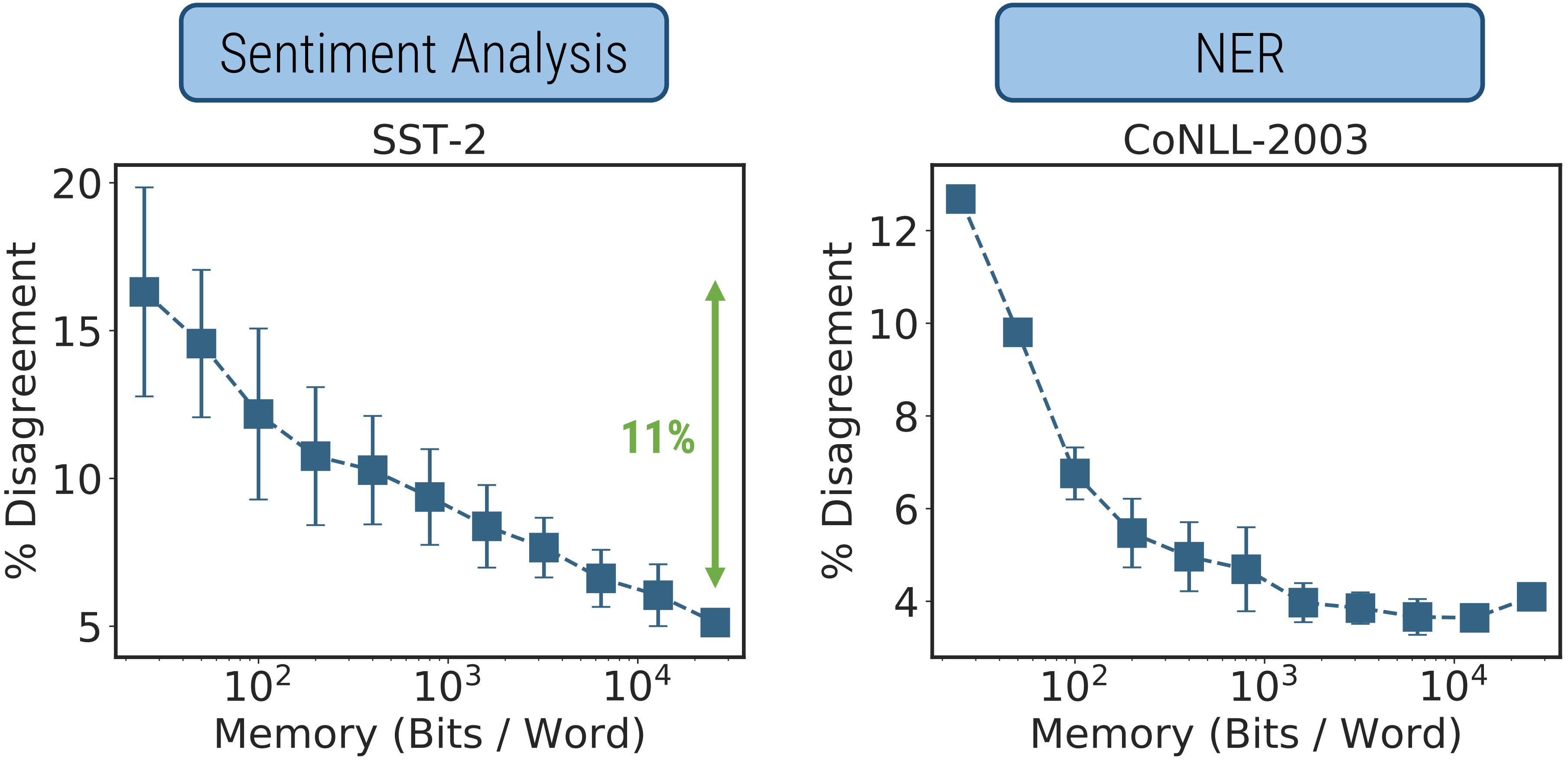
# Impact of Precision




**Precision** **Downstream Instability**

# Stability-Memory Tradeoff




## ⬆ Memory ⬇ 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

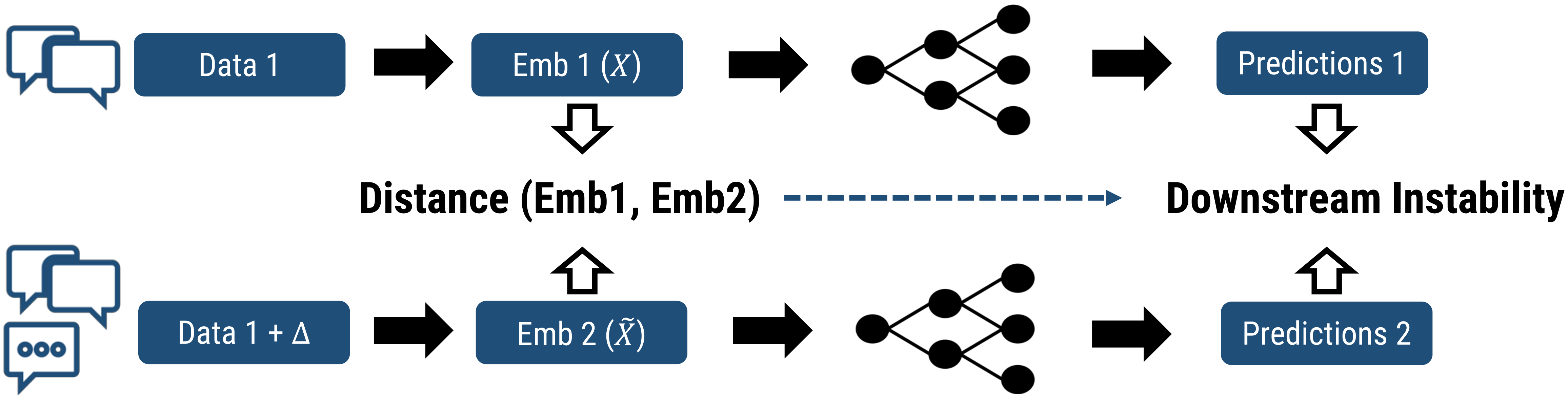
**Q: How can we *theoretically understand* downstream instability?**

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

# 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 $\boldsymbol{X}$ depend on the **left singular vectors** of $\boldsymbol{X}$.[1]

Singular Value Decomposition


[1] May et al. On the downstream performance of compressed word embeddings. NeurIPS, 2019.

# Eigenspace Instability Measure (EIS)

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

For embeddings $\boldsymbol{X}$ and $\widetilde{\boldsymbol{X}}$,

**EIS $(\mathrm{X}, \widetilde{\mathrm{X}})$ = similarity$(\mathrm{U}, \widetilde{\mathrm{U}})$**

- 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 $\mathrm{X}$ and $\widetilde{\mathrm{X}}$.

**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?**

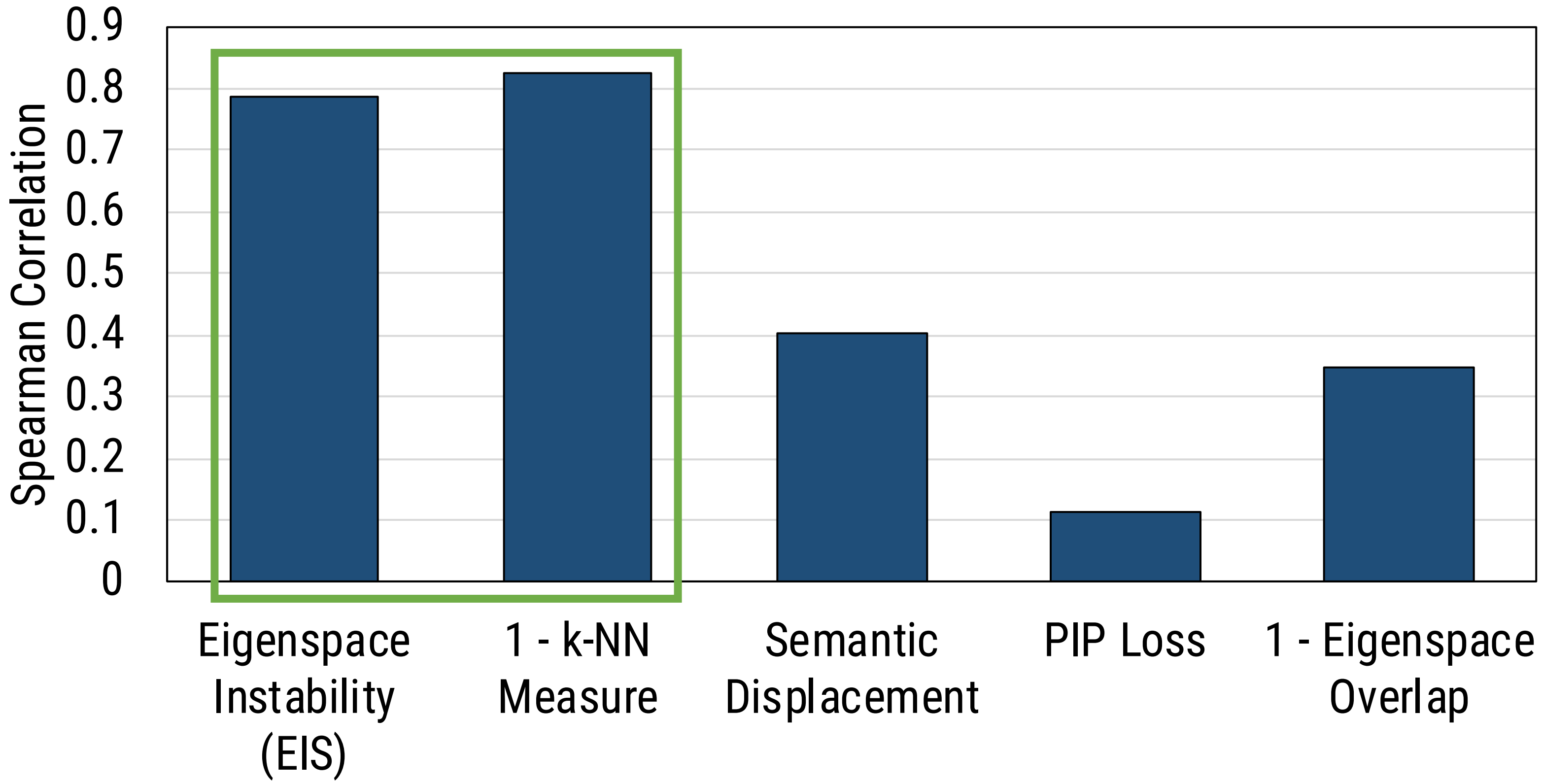
A: using the EIS (or k-NN) measures

# Embedding measure for downstream instability?

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

[1] Hellrich & Hahn, COLING, 2016; [2] Antoniak & Mimno, TACL, 2018; [3] Wendlandt et al., NAACL-HLT, 2018;
[4] Hamilton et al., ACL, 2016; [5] Yin & Shen, NeurIPS, 2018; [6] May et al., NeurIPS, 2019

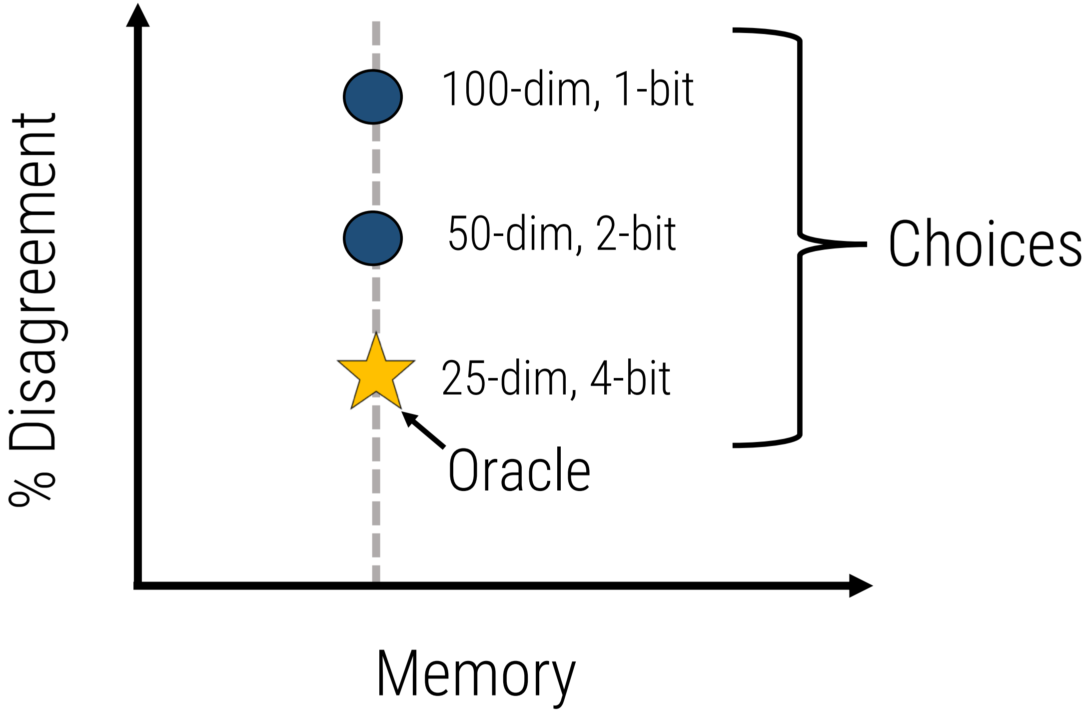
# Correlation with Downstream Instability



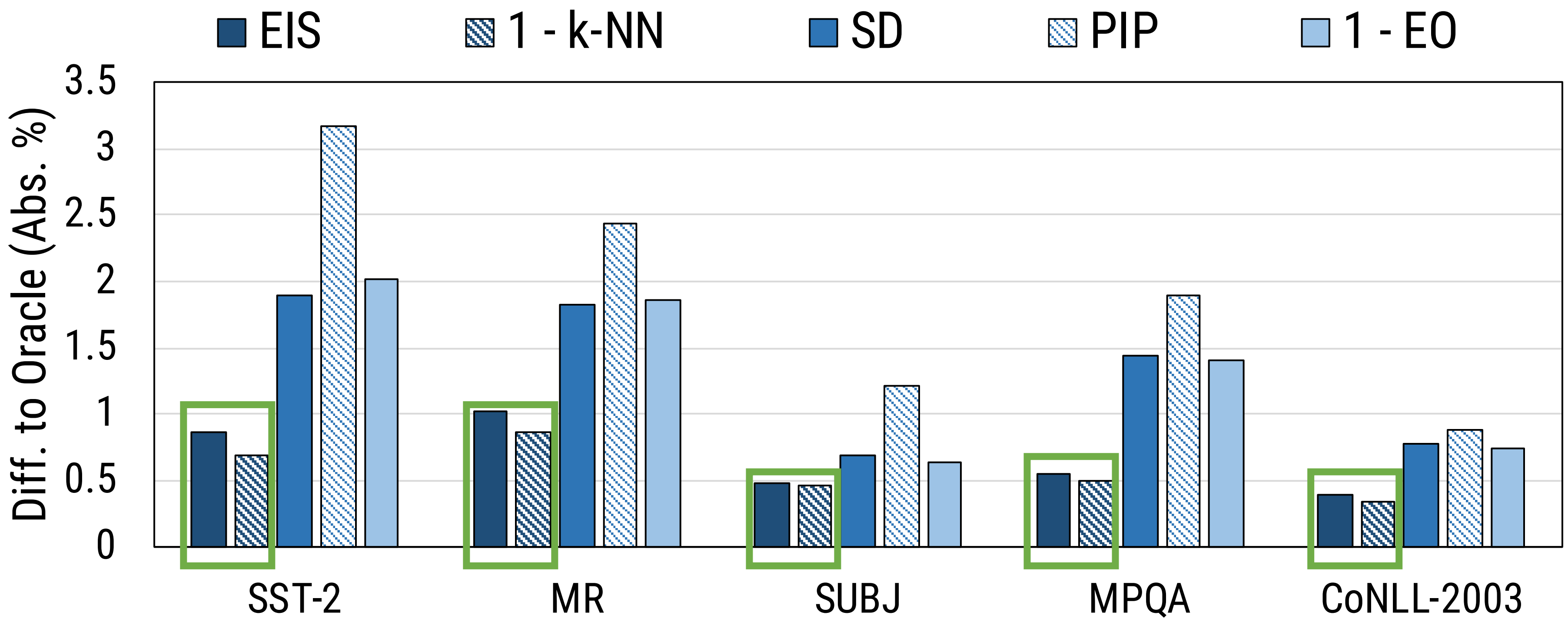

**EIS and k-NN measures strongly correlate with downstream instability.**

# 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



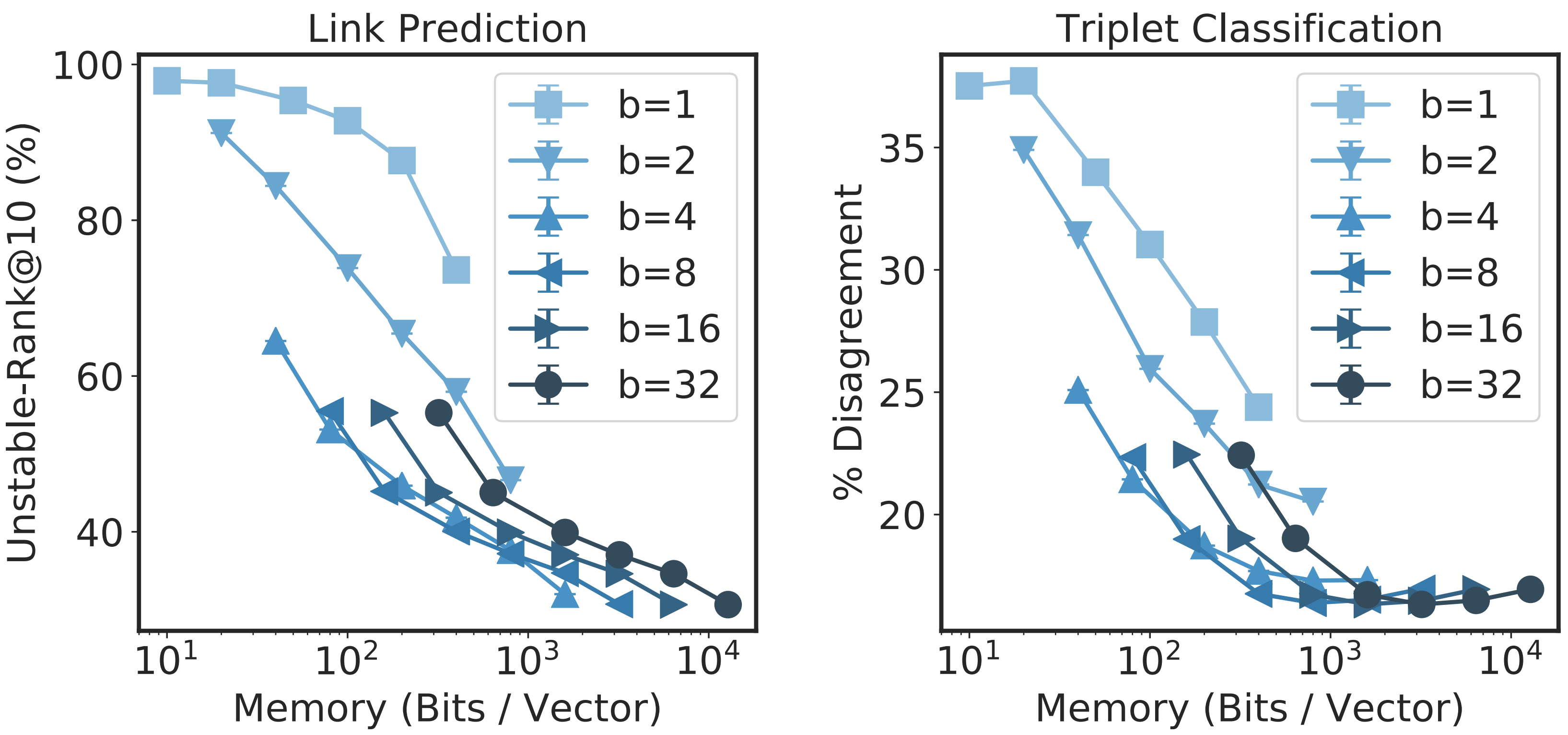
# Selection Task Results

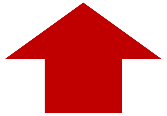



**EIS and k-NN measures outperform other measures as selection criteria.**

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  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:**

http://bit.ly/embstability

**Comments or Questions:**

mleszczy@stanford.edu