
CSAIL

OUR APPROACH

TAILORING OT TO UNREGISTERED SPACES: APPROACH 2

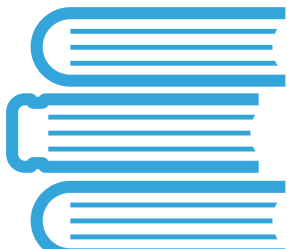
- Word embedding algorithms perform metric recovery [Hashimoto et al., 2016]

- Can we ***"align"*** **the metrics** directly instead?

Semantic Metric Space $(\mathscr{W}, d)$

$$d : \mathscr{W} \times \mathscr{W} \rightarrow \mathbb{R}^{+}$$

Corpus



[Harris, '54; Firth '57]

[Rumelhart + Abrahamson, 1973; Sternberg + Gardner, 1983]

Word2vec, Glove, etc

[Hashimoto et al '16]

Embedding Space $(\mathbb{R}^n, \hat{d})$


dîEnglish


Spanish

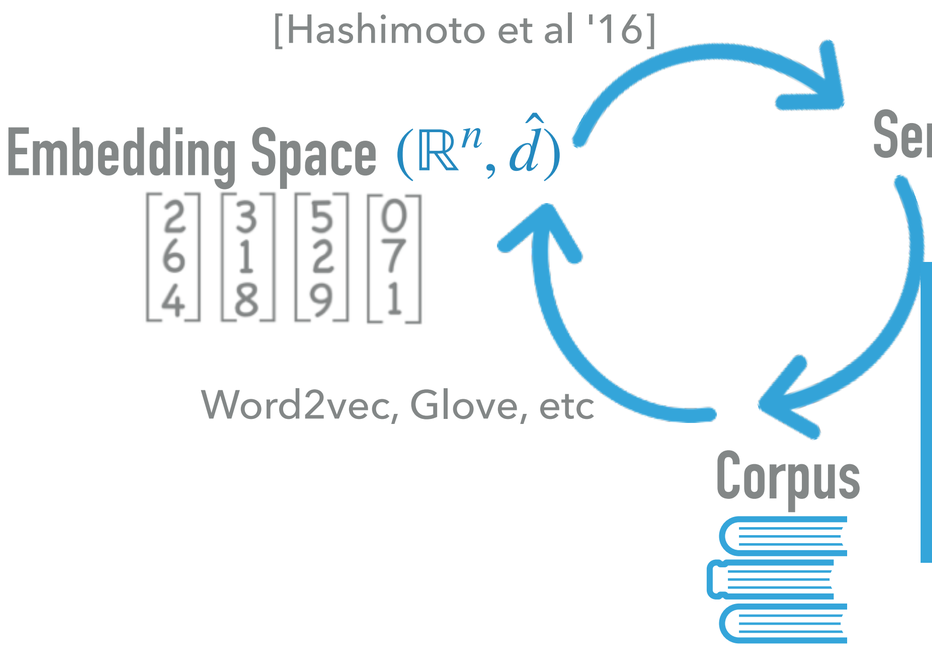
REPRESENTATION-DEPENDENT

REPRESENTATION-INDEPENDENT

MAIN TAKEAWAY: WORD EMBEDDING ALGORITHMS ONLY REALLY CARE ABOUT THE METRIC

TAILORING OT TO UNREGISTERED SPACES: APPROACH 2

- Word embedding algorithms perform metric recovery [Hashimoto et al., 2016]



Semantic Metric Space $(\mathscr{W}, d)$

$d : \mathscr{W} \times \mathscr{W} \rightarrow \mathbb{R}^+$

MAIN TAKEAWAY: WORD EMBEDDING ALGORITHMS ONLY REALLY CARE ABOUT THE METRIC

- Can we ***"align"* the metrics** directly instead? $\hat{d}^{English} \leftrightarrow \hat{d}^{Spanish}$

CSAIL

THE GROMOV-WASSERSTEIN DISTANCE
[Mémoli, 2011; Peyré et al. 2016]