Closed Form Word Embedding Alignment

Sunipa Dev (sunipad@cs.utah.edu)
Safia Hassan
Jeff M Phillips

University of Utah
School of Computing

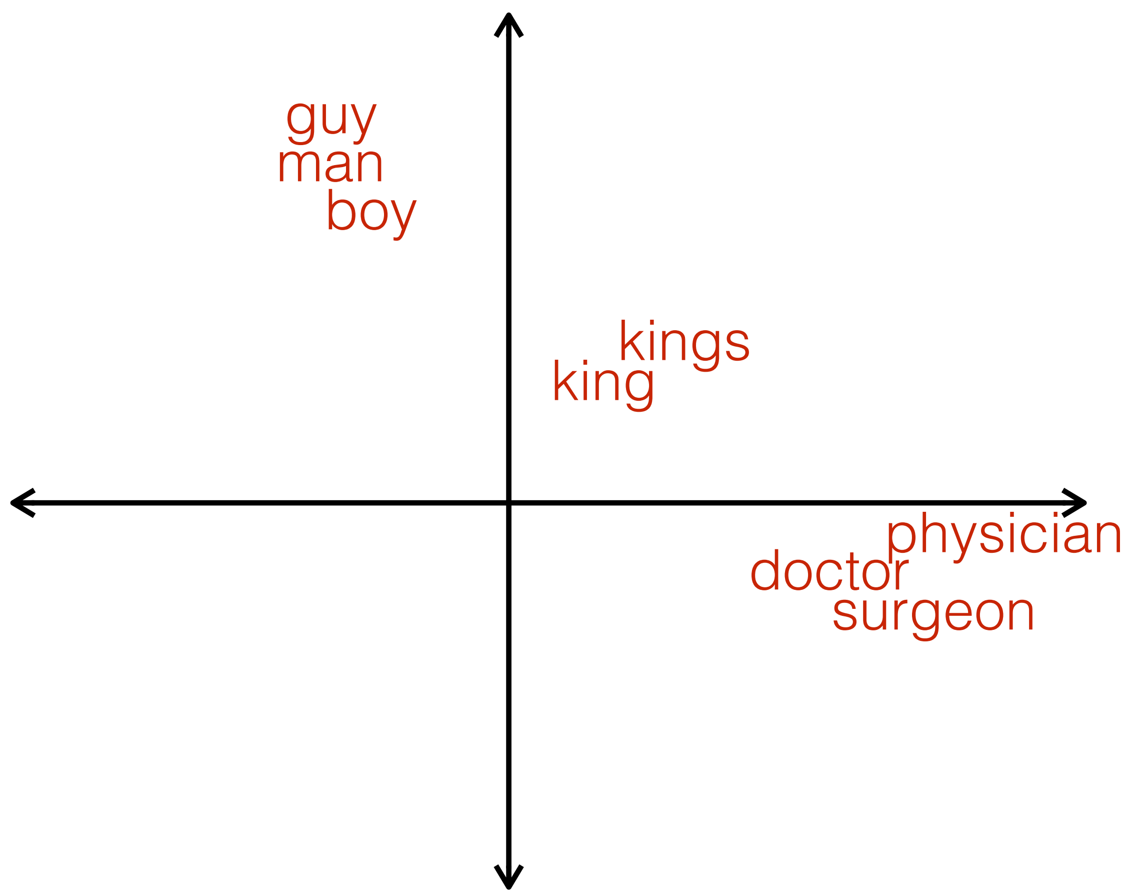
November 10th
ICDM 2019

THE UNIVERSITY OF UTAH®

Closed Form Word Embedding Alignment

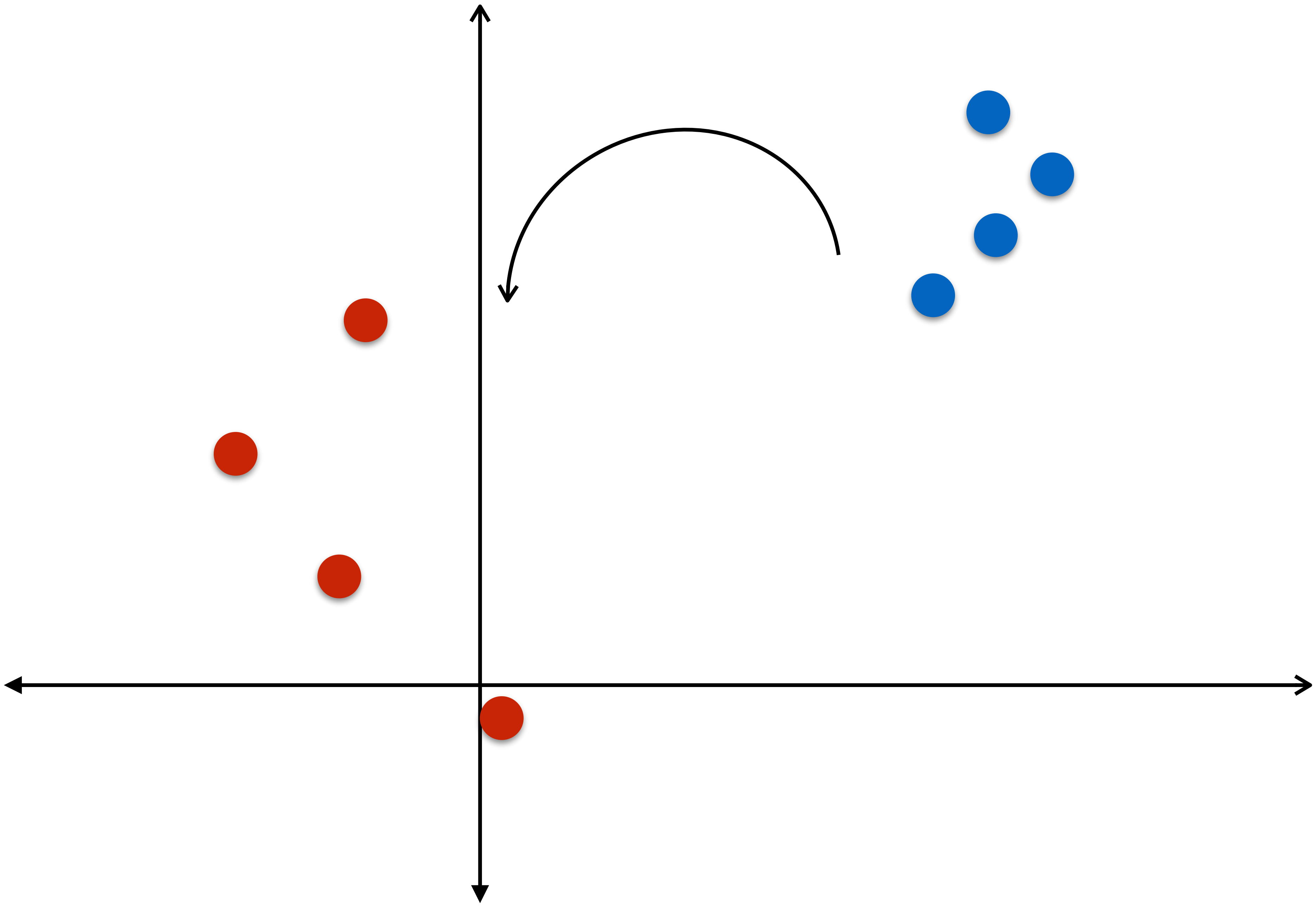
Word2Vec, GloVe, FastText

- distributed representations of word vectors
- low dimensional (about ~300 dimensions)
- contain useful semantic and syntactic information and useful linear relationships



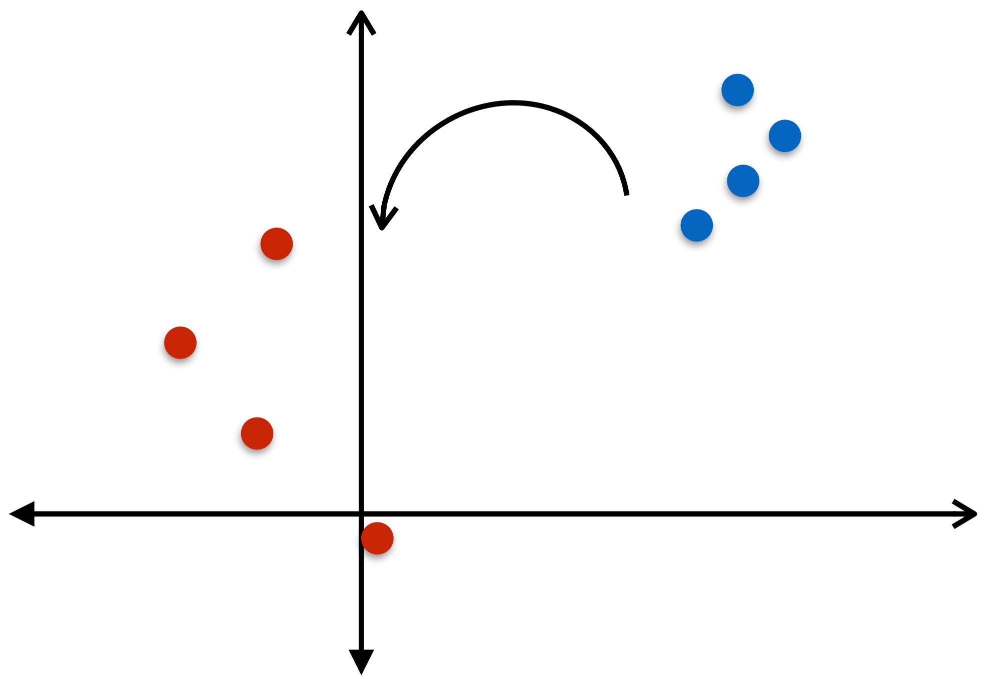
Closed Form Word Embedding Alignment

Given two point sets A and B with correspondence between the points a_i and b_i known, can we align them in space?



Why Align Embeddings?

- In word embeddings :
  - Align phrases, extend pools of synonyms
  - Machine Translation
  - Boosting performance
- In Graph and RDF embeddings to align new nodes



Embedding Alignment by Absolute Orientation

Given two point sets A and B with correspondence between respective points a_i and b_i known,

$$(R^*, t^*, s^*) = \underset{s \in \mathbb{R},\, t \in \mathbb{R}^d,\, R \in \mathbb{SO}(d)}{\operatorname{argmin}} \sum_{i=1}^{n} \|a_i - s(b_i - t)R\|^2$$

Hanson and Norris; Analysis of measurements based on the singular value decomposition. SIAM Journal of Scientific and Statistical Computing,1981

Smith etal; Offline bilingual word vectors, orthogonal transformations and the inverted softmax, ICLR 2017.

Embedding Alignment by Absolute Orientation

$$(R^*, t^*, s^*) = \underset{s \in \mathbb{R}, t \in \mathbb{R}^d, R \in \mathbb{SO}(d)}{\operatorname{argmin}} \sum_{i=1}^{n} \|a_i - s(b_i - t)R\|^2$$

Embedding Alignment by Absolute Orientation

$$(R^*, t^*, s^*) = \underset{s \in \mathbb{R},\, t \in \mathbb{R}^d,\, R \in \mathbb{SO}(d)}{\operatorname{argmin}} \sum_{i=1}^{n} \|a_i - s(b_i - t)R\|^2$$

$$R^* = \underset{R \in \mathbb{SO}(d)}{\operatorname{argmin}} \sum_{i=1}^{n} \|a_i - (b_i R)\|^2.$$ — rotation

Embedding Alignment by Absolute Orientation

$$(R^*, t^*, s^*) = \underset{s\in\mathbb{R},\, t\in\mathbb{R}^d,\, R\in\mathbb{SO}(d)}{\operatorname{argmin}} \sum_{i=1}^{n} \|a_i - s(b_i - t)R\|^2$$

$$R^* = \underset{R\in\mathbb{SO}(d)}{\operatorname{argmin}} \sum_{i=1}^{n} \|a_i - (b_i R)\|^2.$$

— rotation

$$R^* = \underset{R\in\mathbb{SO}(d)}{\operatorname{argmin}} \left(\sum_{i=1}^{n} \|a_i\|^2 - \sum_{i=1}^{n} 2\langle a_i, b_i R\rangle + \sum_{i=1}^{n} \|b_i R\|^2 \right).$$

$$R^* = \underset{R\in\mathbb{SO}(d)}{\operatorname{argmax}} \sum_{i=1}^{n} \langle a_i, b_i R\rangle.$$

Embedding Alignment by Absolute Orientation

$$(R^*, t^*, s^*) = \underset{s\in\mathbb{R},\, t\in\mathbb{R}^d,\, R\in\mathbb{SO}(d)}{\operatorname{argmin}} \sum_{i=1}^{n} \|a_i - s(b_i - t)R\|^2$$

$$R^* = \underset{R\in\mathbb{SO}(d)}{\operatorname{argmin}} \sum_{i=1}^{n} \|a_i - (b_i R)\|^2.$$

— rotation

$$R^* = \underset{R\in\mathbb{SO}(d)}{\operatorname{argmin}} \left(\sum_{i=1}^{n} \|a_i\|^2 - \sum_{i=1}^{n} 2\langle a_i, b_i R\rangle + \sum_{i=1}^{n} \|b_i R\|^2 \right).$$

$$R^* = \underset{R\in\mathbb{SO}(d)}{\operatorname{argmax}} \sum_{i=1}^{n} \langle a_i, b_i R\rangle.$$

$$H = \textstyle\sum_{i=1}^{n} b_i^T a_i$$
$$[U, S, V^T] = \mathsf{svd}(H)$$
$$R = UV^T$$

Embedding Alignment by Absolute Orientation

$$(R^*, t^*, s^*) = \operatorname*{argmin}_{s \in \mathbb{R},\, t \in \mathbb{R}^d,\, R \in \mathbb{SO}(d)} \sum_{i=1}^{n} \| a_i - s(b_i - t)R \|^2$$

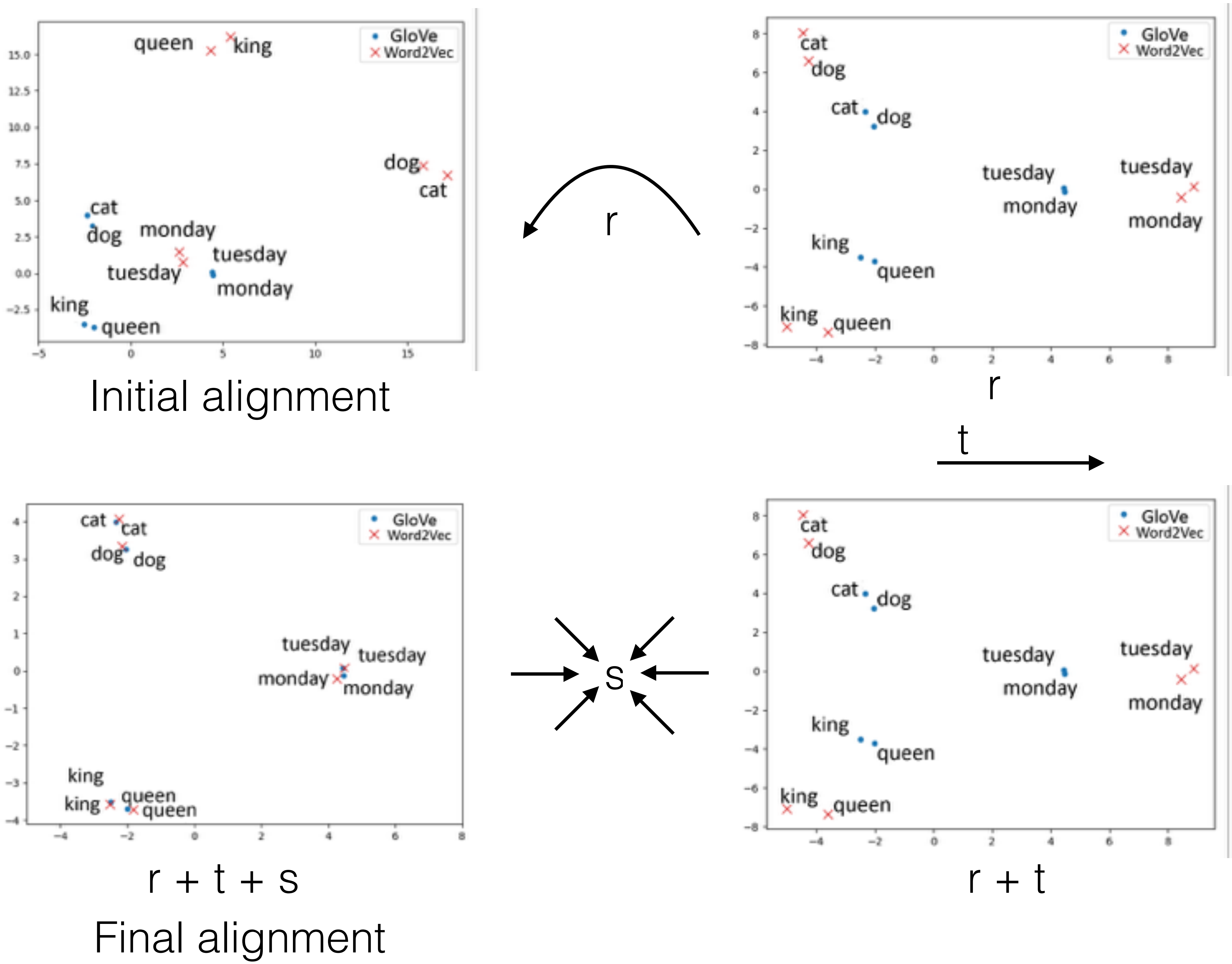
$$t^* = \frac{1}{n} \sum_{i=1}^{n} b_i - \frac{1}{n} \sum_{i=1}^{n} a_i$$ — translation

Embedding Alignment by Absolute Orientation

$$(R^*, t^*, s^*) = \underset{s\in\mathbb{R},\, t\in\mathbb{R}^d,\, R\in\mathbb{SO}(d)}{\operatorname{argmin}} \sum_{i=1}^{n} \|a_i - s(b_i - t)R\|^2$$

$$t^* = \frac{1}{n}\sum_{i=1}^n b_i - \frac{1}{n}\sum_{i=1}^n a_i$$ — translation

$$s^* = \frac{\sum_{i=1}^n \langle a_i, b_i\rangle}{\sum_{i=1}^n \|b_i\|^2} = \frac{\sum_{i=1}^n \langle a_i, b_i\rangle}{\|B\|_F^2}$$ — scaling



GloVe
Word2Vec
queen
king
dog
cat
monday
tuesday
Initial alignment
r
t
r + t
s
r + t + s
Final alignment

Related Techniques

For points a_i and b_i in embeddings A and B respectively, and parameter γ,

$$\underset{M \in \mathbb{R}^{d \times d}}{\operatorname{argmin}} \sum_{i=1}^{n} \|a_i - b_i M\|^2 + \gamma \|M\|_F^2.$$

Bollegela *et al ; Learning linear transformations between counting-based and prediction-based word embeddings; PloS ONE, 12(9):3370–3374, 2017*

Related Techniques

For embeddings A and B respectively, their correspondence matrix C and appropriately chosen λ,

$$M_A = \underset{M \in \mathbb{R}^{n \times n}}{\operatorname{argmin}} \frac{1}{2}\|A - AM\|_F^2 + \lambda\|M\|.$$

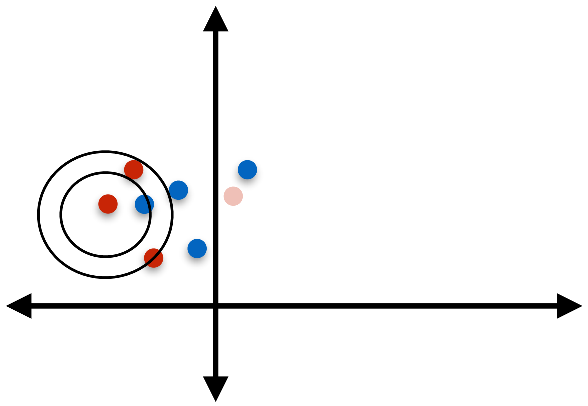
$$M_B = \underset{M \in \mathbb{R}^{n \times n}}{\operatorname{argmin}} \frac{1}{2}\|B - BM\|_F^2 + \lambda\|M\|.$$

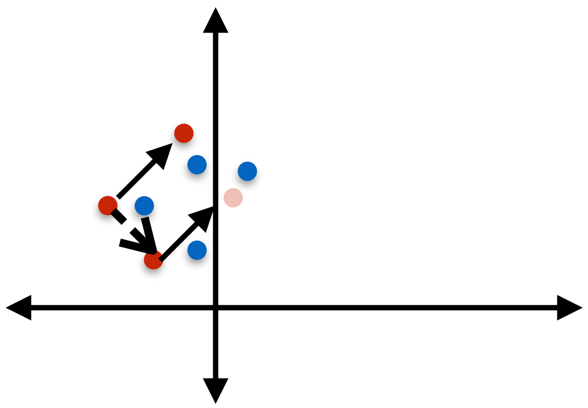
$$M = \begin{bmatrix} M_A & 0 \\ 0 & M_B \end{bmatrix}$$

Sahin *et al ; Consistent Alignment of Word Embedding Models. ArXiv e-prints, February 2017*

Related Techniques

Test Sets	LRA	Affine Transformation	Absolute Orientation (10 K words)	Absolute Orientation (100 K words)
RG	0.701	0.301	0.728	0.818
WSIM	0.616	0.269	0.612	0.618
SYN	0.719	0.412	0.722	0.766
SEM	0.327	0.126	0.340	0.343





Closed Form Word Embedding Alignment

- Resilience of word embeddings

- Relative spatial distribution of words the same across embeddings

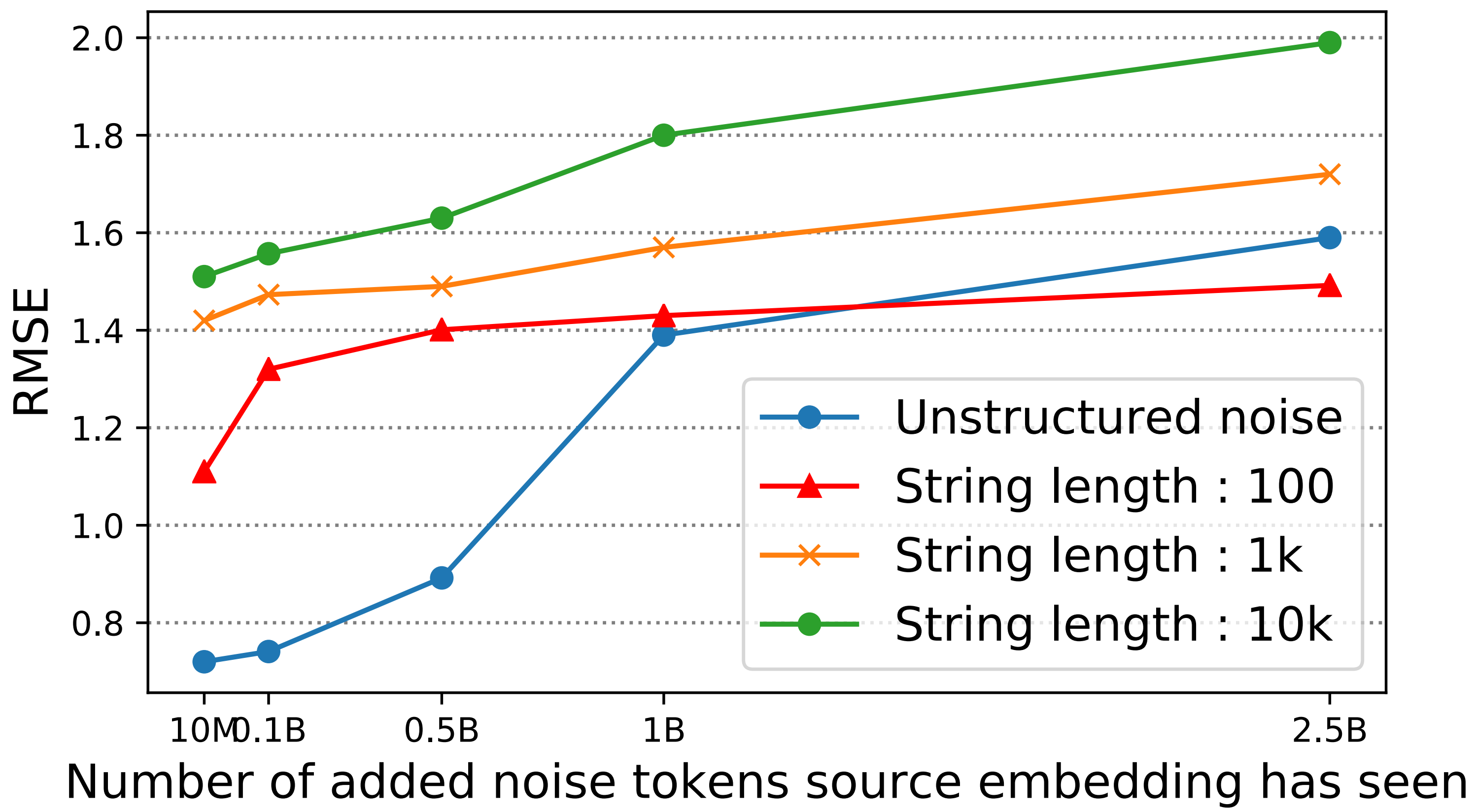
- Frequency implies stability

Closed Form Word Embedding Alignment

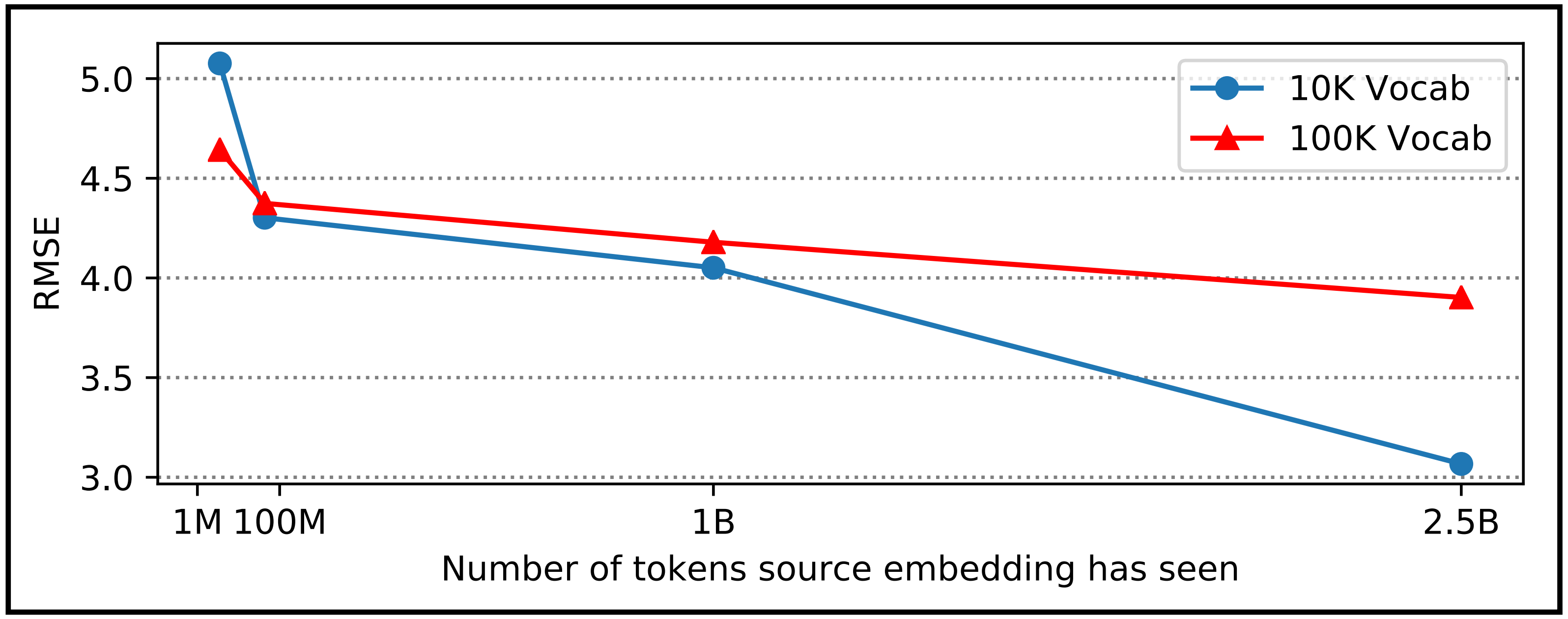
- Resilience of word embeddings
- Relative spatial distribution of words the same across embeddings
- Frequency implies stability

Wendlandt etal; "Factors Influencing the Surprising Instability of Word Embeddings." In Proceedings of NAACL : Human Language Technologies, 2018

Resilience of Word Embeddings



Resilience of Word Embeddings

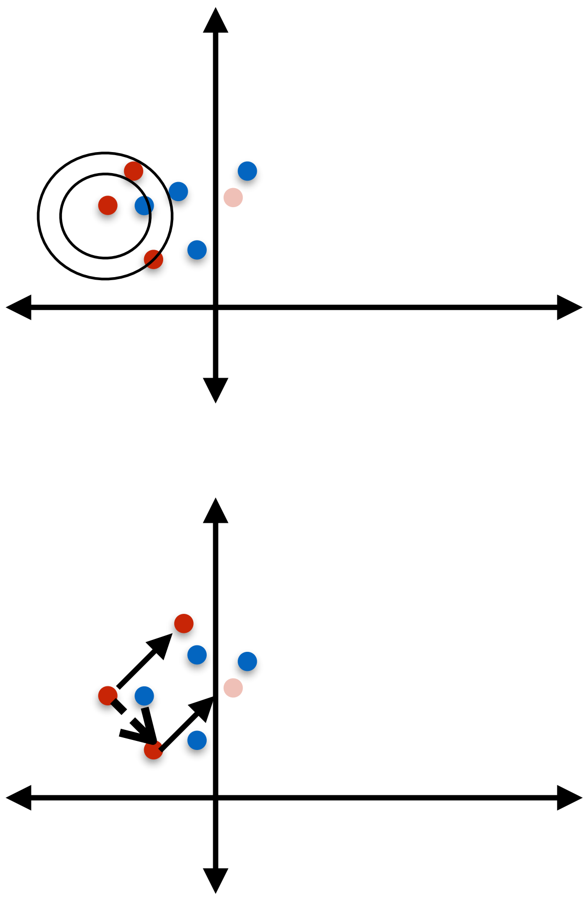


Closed Form Word Embedding Alignment

- Resilience of word embeddings
- Relative spatial distribution of words the same across embeddings
- Frequency implies stability

Relative Spatial Distribution of Word Vectors

Test Sets	GloVe(Wiki)	word2vec(Wiki)	Original	+r +t + s
RG	0.614	0.696	0.041	0.584
WSIM	0.623	0.659	0.064	0.625
SYN	0.587	0.582	0.0	0.501
SEM	0.691	0.722	0.0	0.624

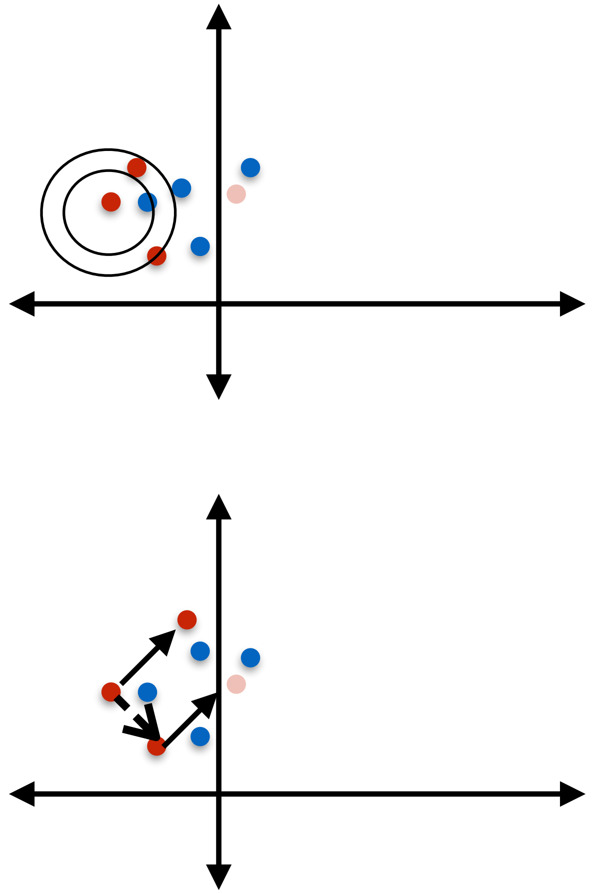


(higher scores = better performance)

Relative Spatial Distribution of Word Vectors

Test Sets	GloVe(Wiki)	GloVe(CC)	Original	+r
RG	0.614	0.696	0.363	0.616
WSIM	0.623	0.659	0.017	0.618
SYN	0.587	0.582	0.0	0.566
SEM	0.691	0.722	0.0	0.676

(higher scores = better performance)



Closed Form Word Embedding Alignment

- Resilience of word embeddings
- Relative spatial distribution of words the same across embeddings
- Frequency implies stability

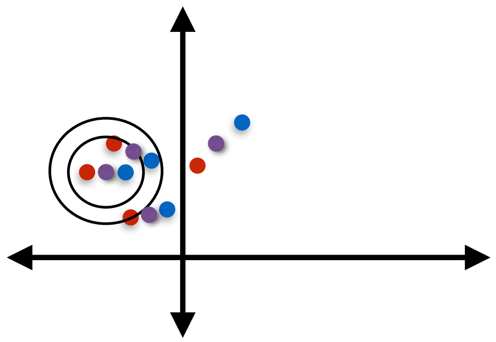
Applications

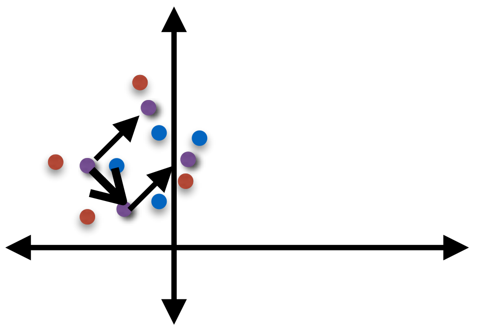
- Boosting via ensembles
- Translation between languages
- Adding new word vectors inexpensively

Applications

- Boosting via ensembles
- Translation between languages
- Adding new word vectors inexpensively

Test Sets	G(Wiki)	W(Wiki)	G(Wiki) + W(Wiki)
RG	0.614	0.696	0.715
WSIM	0.623	0.659	0.697
SYN	0.587	0.582	0.594
SEM	0.691	0.722	0.757





Applications

- Boosting via ensembles
- Translation between languages
- Adding new word vectors inexpensively

Word	Neighbors before alignment	Neighbors after alignment
woman	her, young, man, girl, mother	her, girl, **mujer**, mother, man
week	month, day, year, monday, time	days, **semana**, year, day, month

Applications

- Boosting via ensembles
- Translation between languages
- Adding new word vectors inexpensively

	Top 10 accuracy (before)	Top 10 accuracy (after)
EN - ES	0.054	0.848
EN - FR	0.0	0.701

Applications

- Boosting via ensembles
- Translation between languages
- Adding new word vectors inexpensively

Thank You

sunipad@cs.utah.edu
github.com/sunipa/Abs-Orientation
arxiv.org/abs/1806.01330