

Parameter-free Sentence Embedding via Orthogonal Basis

Ziyi Yang¹, Chenguang Zhu² and Weizhu Chen³

¹Stanford University

²Microsoft Speech and Dialogue Research Group

³Microsoft Dynamics 365 AI



Motivation

What is a sentence, in a mathematical sense?

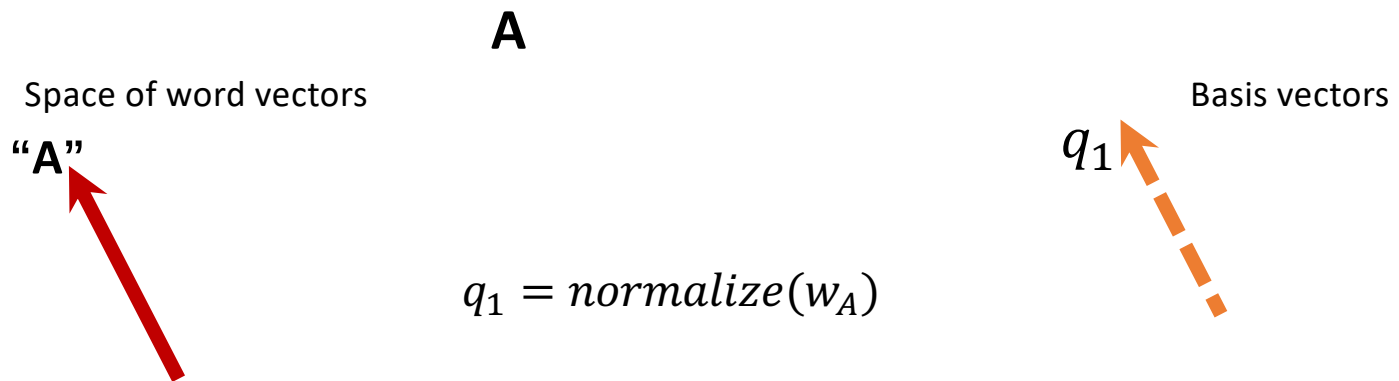
Motivation

A man is dancing

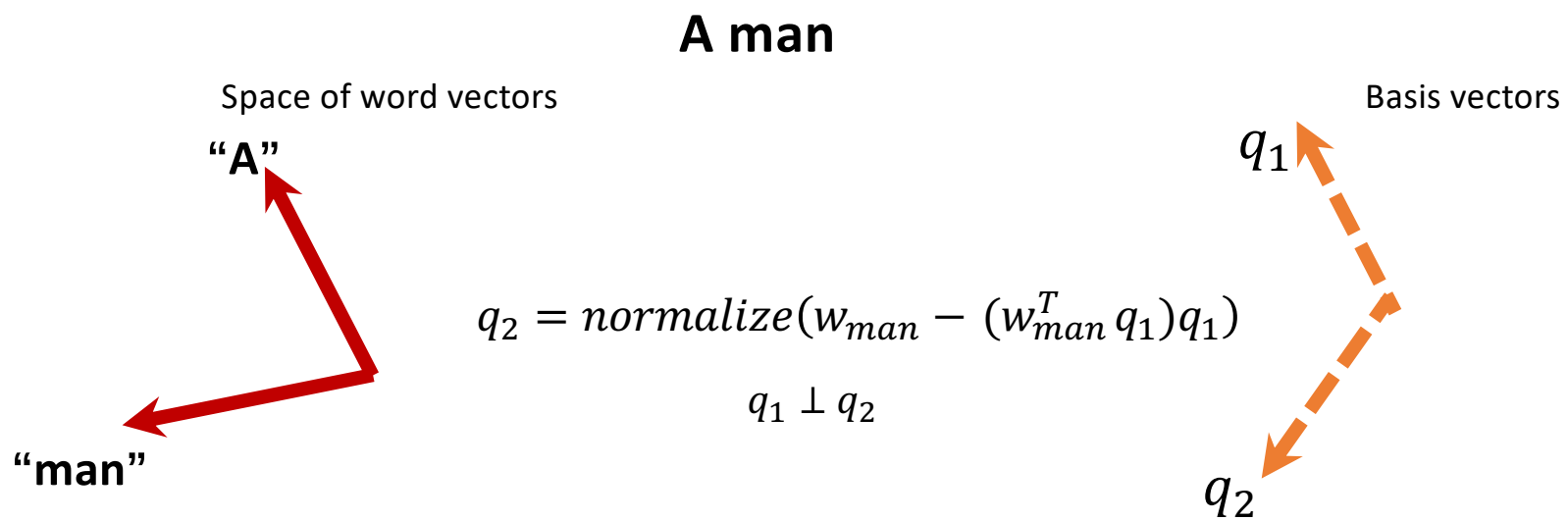
Space of word vectors

Basis vectors

Motivation

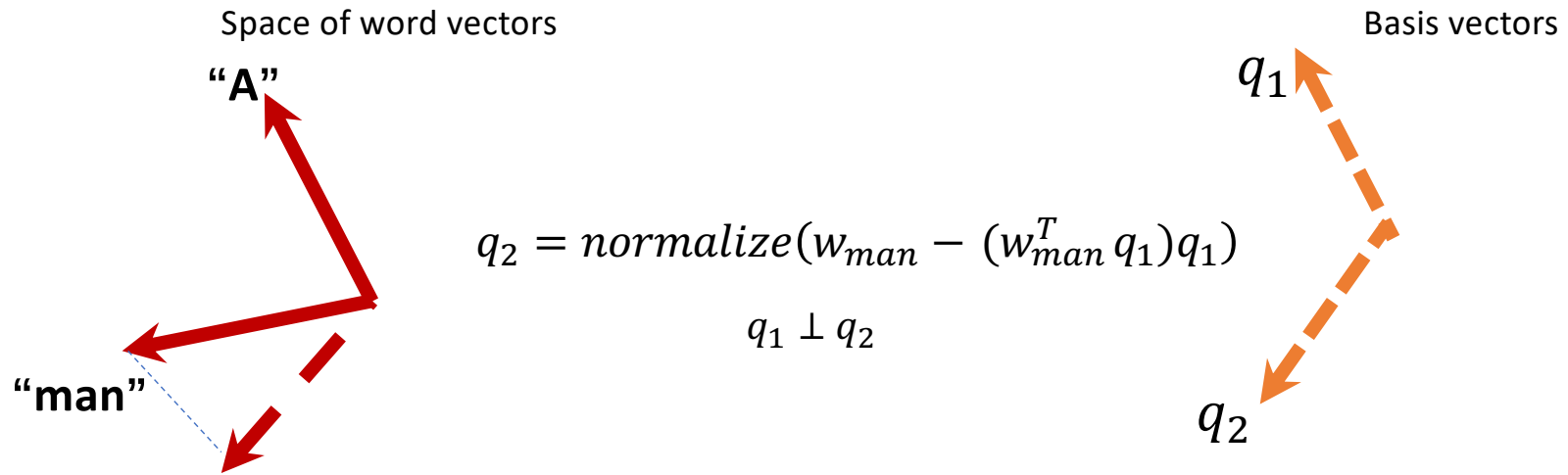


Motivation



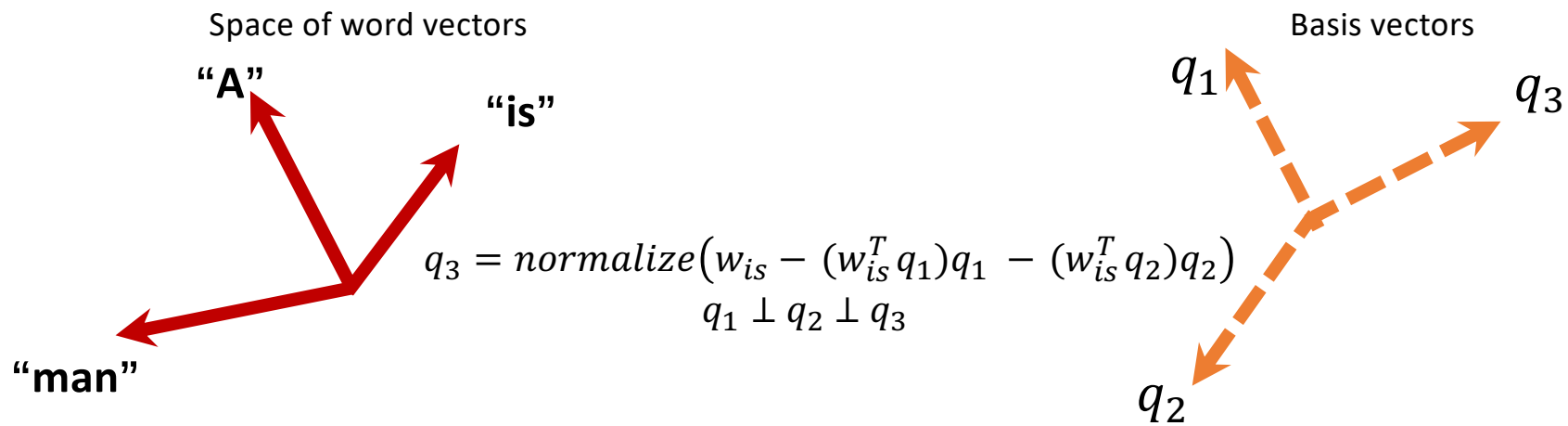
Motivation

A man



Motivation

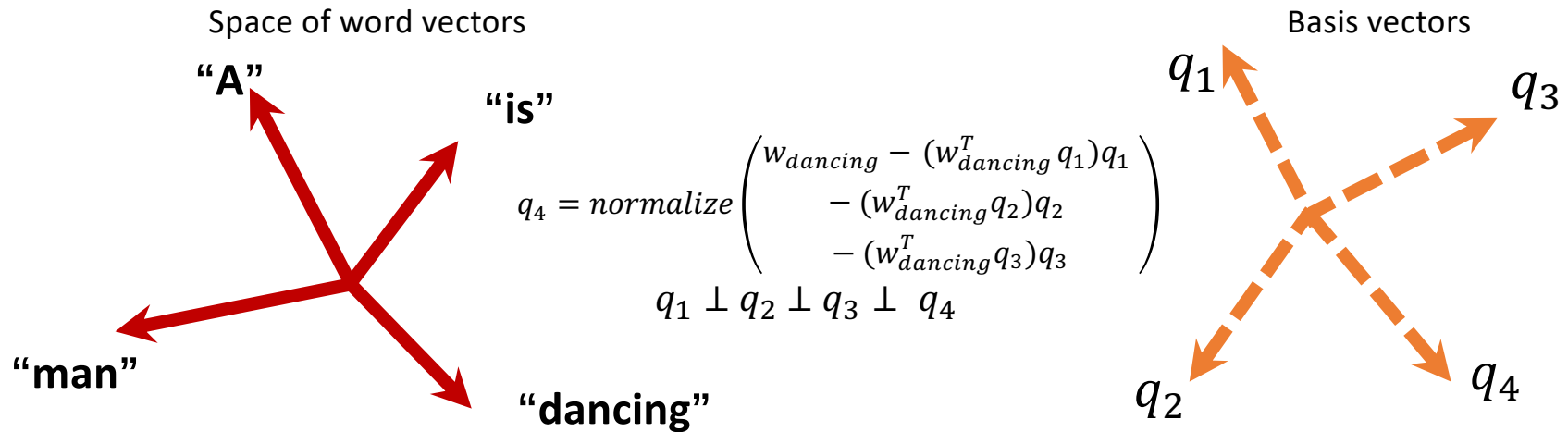
A man is



Motivation

- Sentence is a subspace spanned by its word vectors
- Each word may bring in a new direction (semantic meaning) q_i

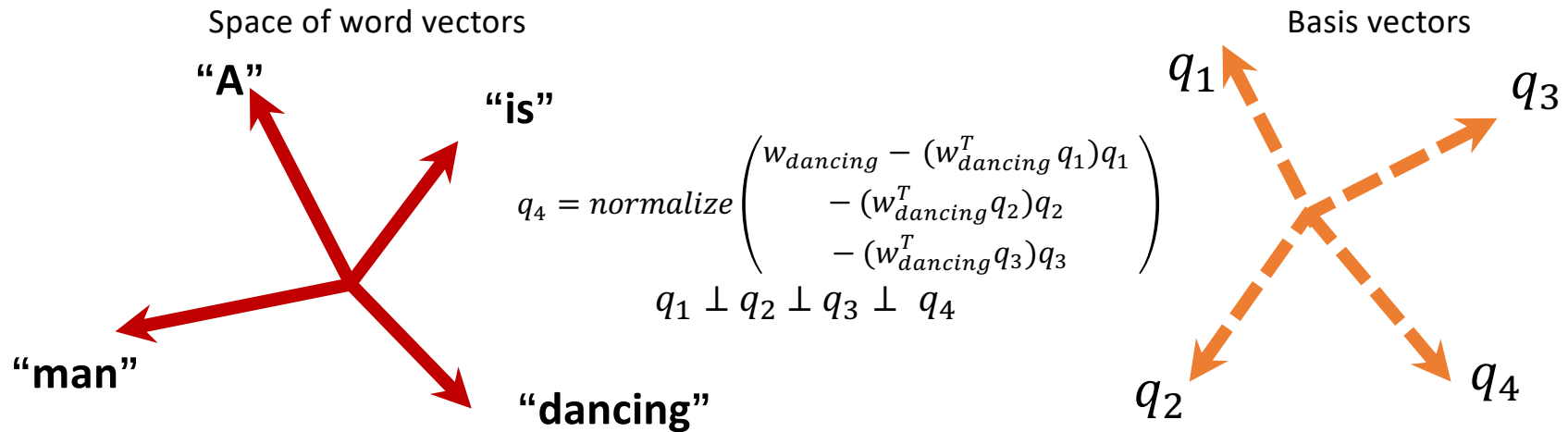
A man is dancing



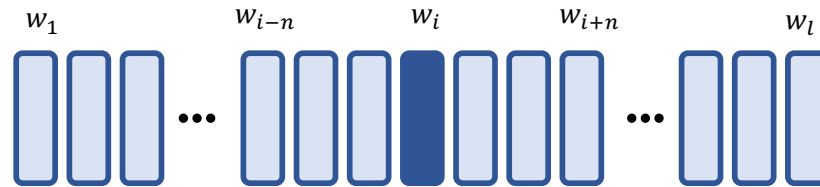
QR factorization/Gram Schmidt Process

- An algorithm to find basis vectors in the subspace
- $S = [w_A, w_{man}, w_{is}, w_{dancing}] = QR, Q = [q_1, q_2, q_3, q_4]$

A man is dancing

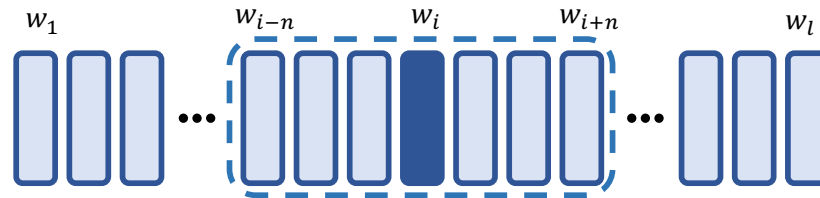


Quantify the new semantic meaning q_i

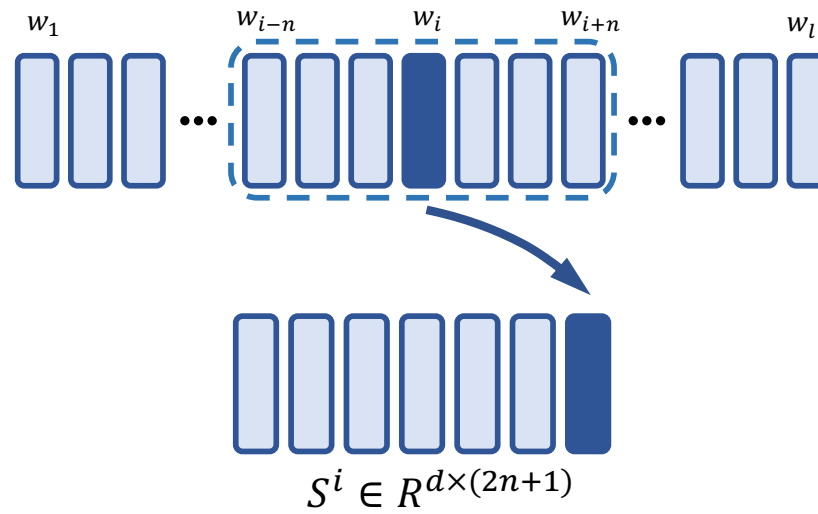


Quantify the new semantic meaning q_i

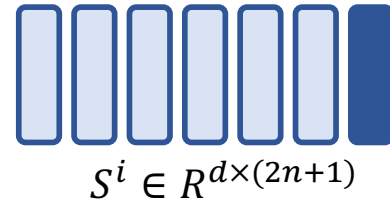
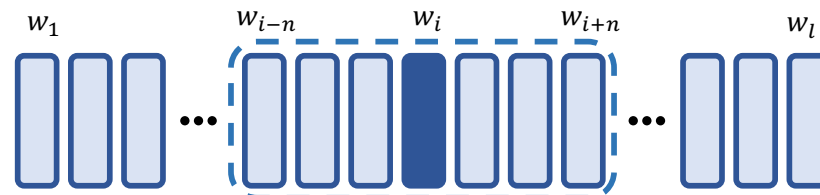
- Contextual window
- Look at both n preceding and n following words



Quantify the new semantic meaning q_i

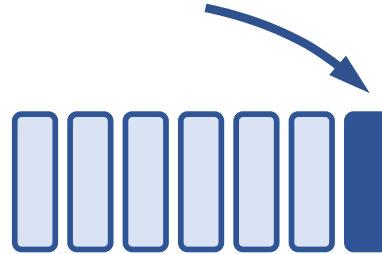
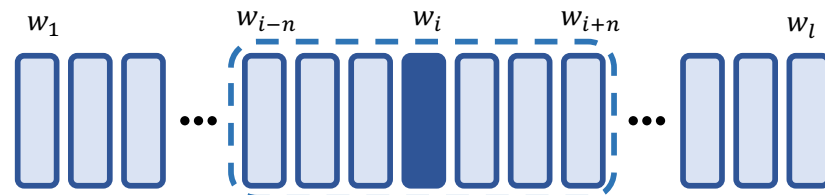


Quantify the new semantic meaning q_i



A dark blue vertical rectangle is followed by an equals sign, then a light red vertical rectangle with the number 0.2 to its left, a minus sign, a light red vertical rectangle with the number 0.4 to its left, an ellipsis, a plus sign, and a dark red vertical rectangle with the number 0.1 to its left. Below this is the equation $S^i = QR$.

Quantify the new semantic meaning q_i



$$S^i \in R^{d \times (2n+1)}$$

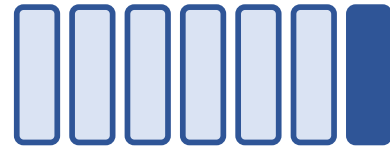
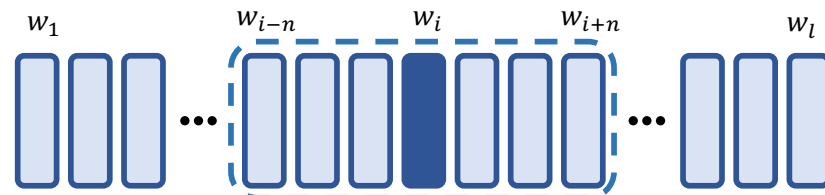
The dark blue bar is equated to a linear combination of red bars: $= 0.2$ (light red bar) $- 0.4$ (medium red bar) $\dots + 0.1$ (dark red bar). The dark red bar is enclosed in a dashed blue box and labeled q_i . Below the equation is the matrix equation $S^i = QR$.

$$S^i = QR$$

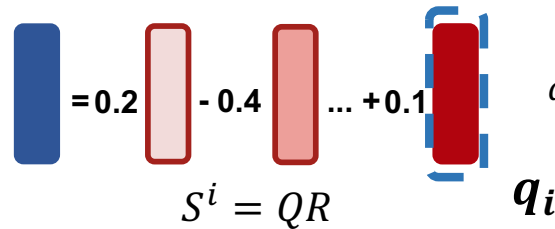
Three-level Weights

- In the ***single word w_i*** , is q_i dominant?
- In w_i 's *context*, is q_i important?
- In the *sentence corpus*, is q_i unique?

In the single word w_i : Word-wise weight α_w



$$S^i \in R^{d \times (2n+1)}$$



$$\alpha_w = \exp\left(\frac{0.1}{\| [0.2, -0.4, \dots, 0.1] \|_2}\right)$$

Three-level Weights

- In the *single word* w_i , is q_i dominant?
- In **w_i 's context**, is q_i important?
- In the *sentence corpus*, is q_i unique?

In w_i 's context: Contextual-wise weight α_c

- Want to know if q_i is important in S^i

In w_i 's context: Contextual-wise weight α_c

- Want to know if q_i is important in S^i
- Need to decide which directions are important in S^i

In w_i 's context: Contextual-wise weight α_c

- Want to know if q_i is important in S^i
- Need to decide which directions are important in S^i
- Singular Value Decomposition (SVD) comes to help!

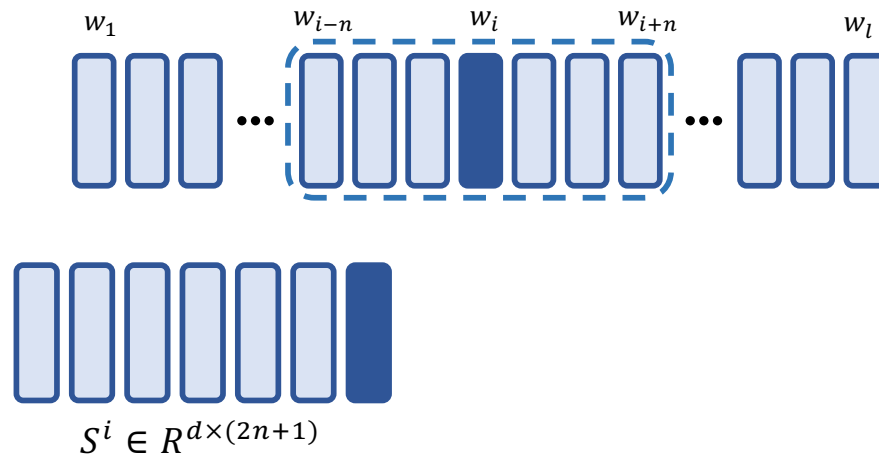
In w_i 's context: Contextual-wise weight α_c

- Want to know if q_i is important in S^i
- Need to decide which directions are important in S^i
- Singular Value Decomposition (SVD) comes to help!

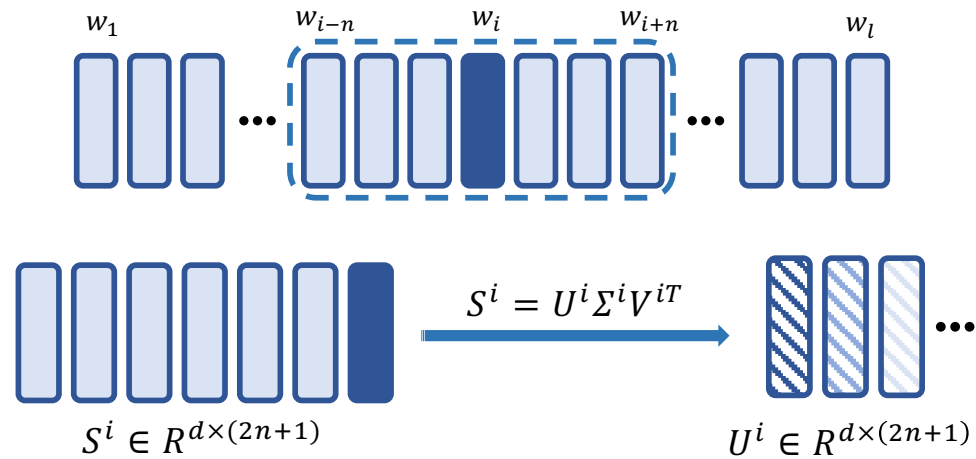
$$S^i = U^i \Sigma^i V^{iT}$$

U^i columns are basis vectors of S^i , and singular values (diag(Σ^i)) are importance scores

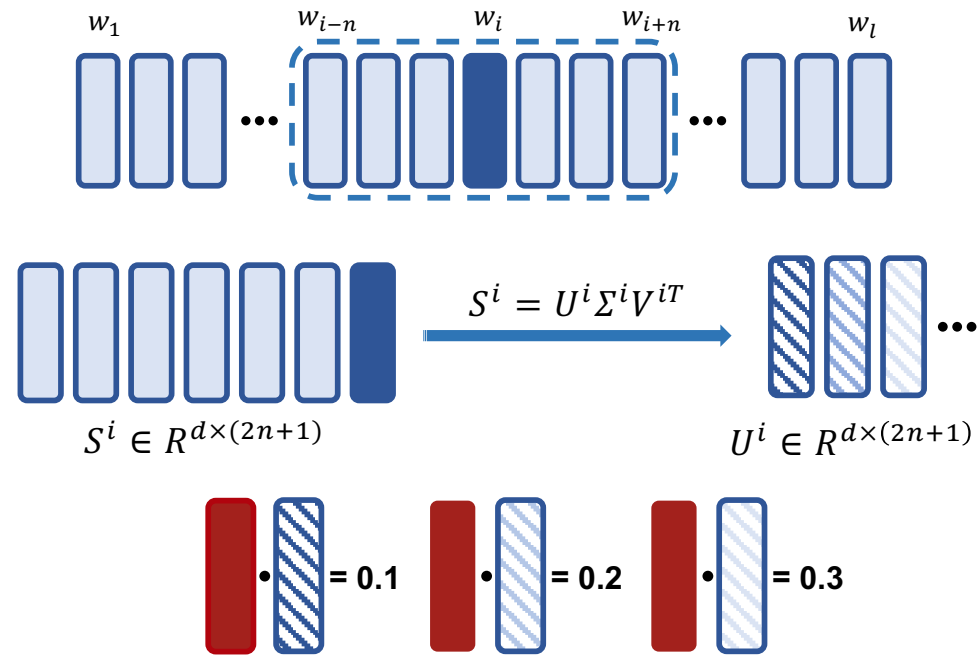
In w_i 's context: Contextual-wise weight α_c



In w_i 's context: Contextual-wise weight α_c

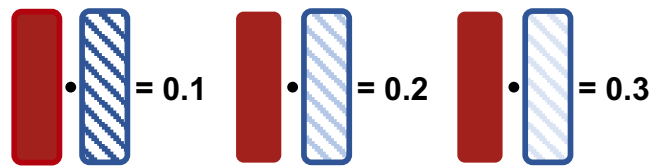
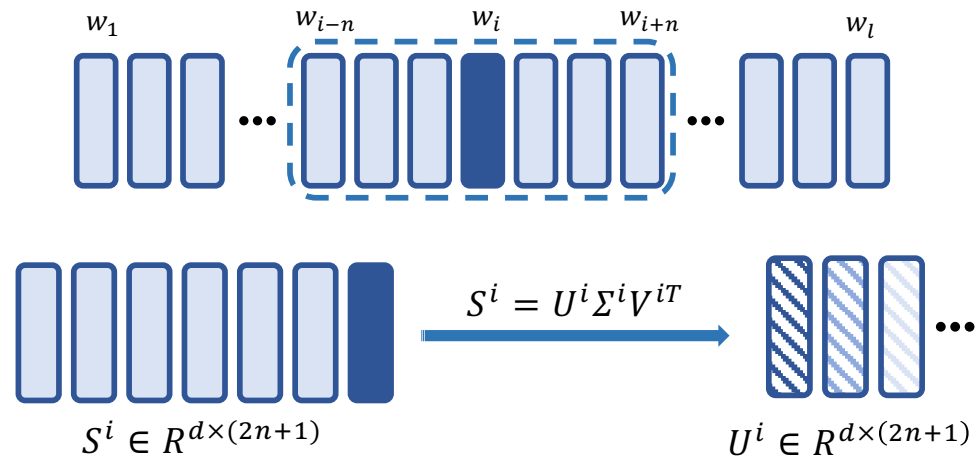


In w_i 's context: Contextual-wise weight α_c



q_i

In w_i 's context: Contextual-wise weight α_c



$$\alpha_c = \frac{\| [0.1\sigma(S^i)_1, 0.2\sigma(S^i)_2, 0.1\sigma(S^i)_3, \dots] \|_2}{2n+1}$$



Three-level Weights

- In the *single word* w_i , is q_i dominant?
- In w_i 's *context*, is q_i important?
- In the ***sentence corpus***, is q_i unique?

In the sentence corpus: Corpus-wise weight α_p

- Need to decide what directions are common in corpus
(s_1, s_2, \dots, s_n)

In the sentence corpus: Corpus-wise weight α_p

- Need to decide what directions are common in corpus
(s_1, s_2, \dots, s_n)
- Encode each sentence $s_k = [w_1, w_2, \dots, w_l]$ into a coarse embedding c_k

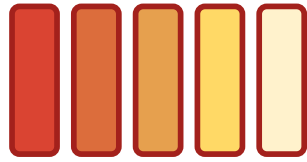
In the sentence corpus: Corpus-wise weight α_p

- Need to decide what directions are common in corpus
(s_1, s_2, \dots, s_n)
- Encode each sentence $s_k = [w_1, w_2, \dots, w_l]$ into a coarse embedding c_k
- $s_k = U\Sigma V^T, c_k = U\text{diag}(\Sigma^3)$

In the sentence corpus: Corpus-wise weight α_p

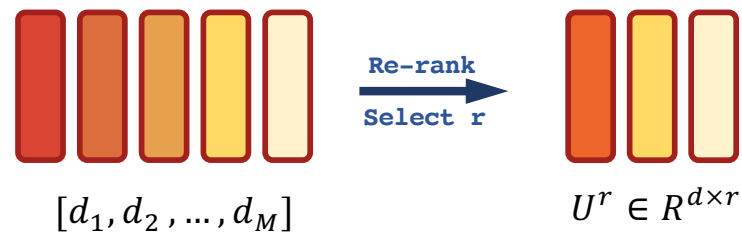
- Need to decide what directions are common in corpus (s_1, s_2, \dots, s_n)
- Encode each sentence $s_k = [w_1, w_2, \dots, w_l]$ into a coarse embedding c_k
- $s_k = U\Sigma V^T, c_k = U\text{diag}(\Sigma^3)$
- Build a corpus matrix $X = [c_1, c_2, \dots, c_n]$, compute top M singular vectors $[d_1, d_2, \dots, d_M]$

In the sentence corpus: Corpus-wise weight α_p



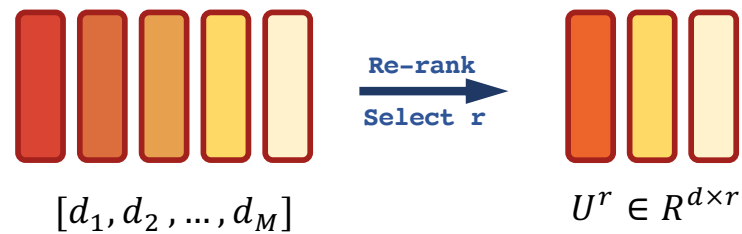
$[d_1, d_2, \dots, d_M]$

In the sentence corpus: Corpus-wise weight α_p

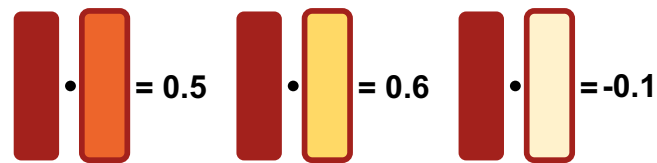


For each sentence c_k , $[d_1, d_2, \dots, d_M]$ are re-ranked by $\sigma_i \|c_k^T d_i\|_2$

In the sentence corpus: Corpus-wise weight α_p

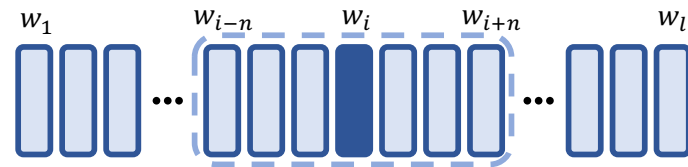


For each sentence c_k , $[d_1, d_2, \dots, d_M]$ are re-ranked by $\sigma_i \|c_k^T d_i\|_2$

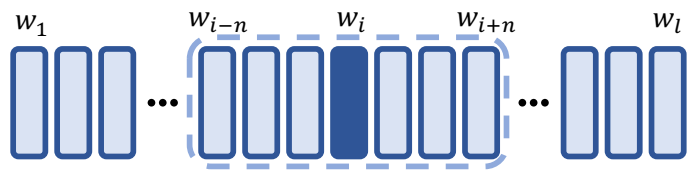


$$\alpha_p = \exp(-\|[0.5, 0.6, -0.1, \dots]\|_2/r)$$

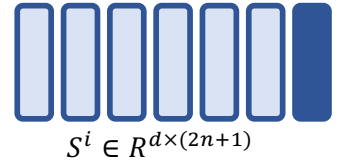
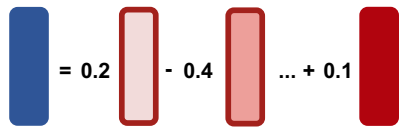
Geometric Embedding (GEM)



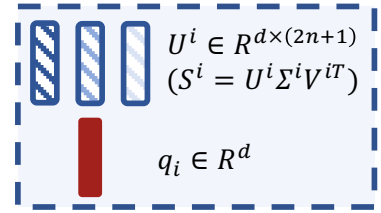
Geometric Embedding (GEM)



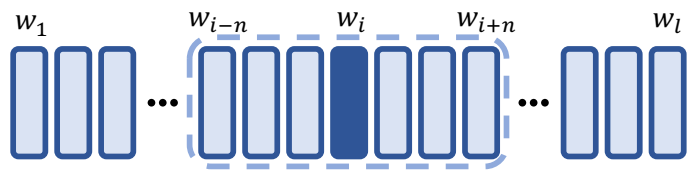
$$\alpha_w = \exp\left(\frac{0.1}{\|[0.2, -0.4, \dots, 0.1]\|_2}\right)$$



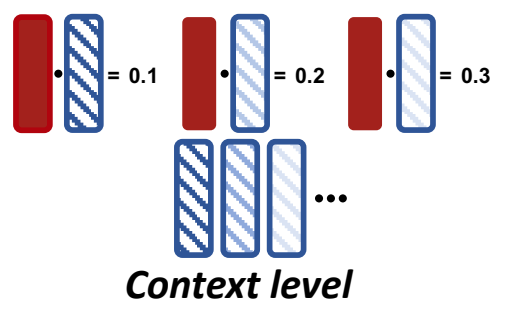
single word level



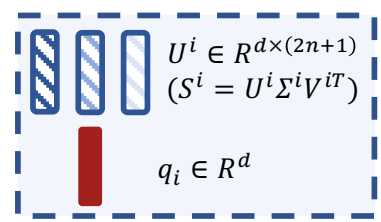
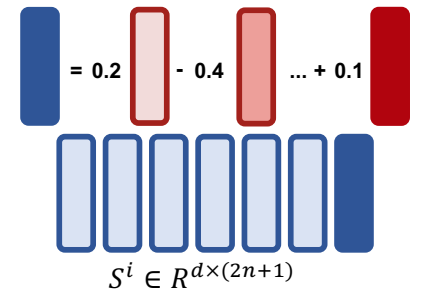
Geometric Embedding (GEM)



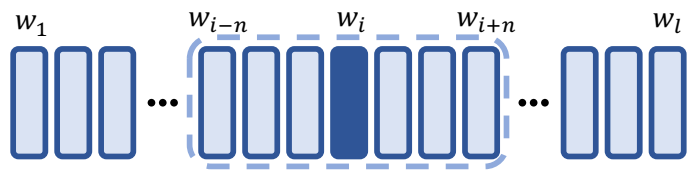
$$\alpha_c = \frac{\| [0.1\sigma(S^i)_1, 0.2\sigma(S^i)_2, 0.1\sigma(S^i)_3, \dots] \|_2}{2n+1}$$



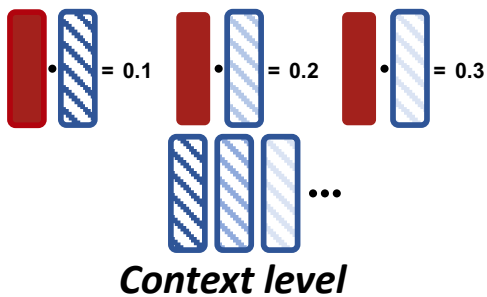
$$\alpha_w = \exp\left(\frac{0.1}{\| [0.2, -0.4, \dots, 0.1] \|_2}\right)$$



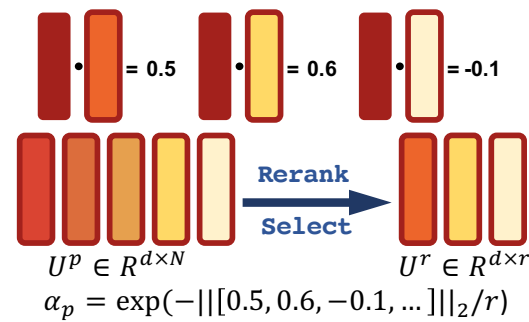
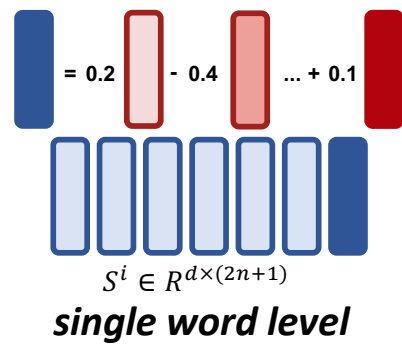
Geometric Embedding (GEM)



$$\alpha_c = \frac{\| [0.1\sigma(S^i)_1, 0.2\sigma(S^i)_2, 0.1\sigma(S^i)_3, \dots] \|_2}{2n + 1}$$

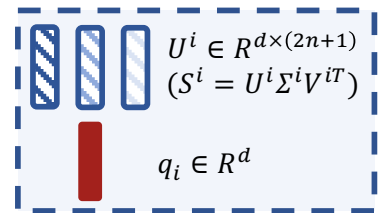


$$\alpha_w = \exp\left(\frac{0.1}{\|[0.2, -0.4, \dots, 0.1]\|_2}\right)$$

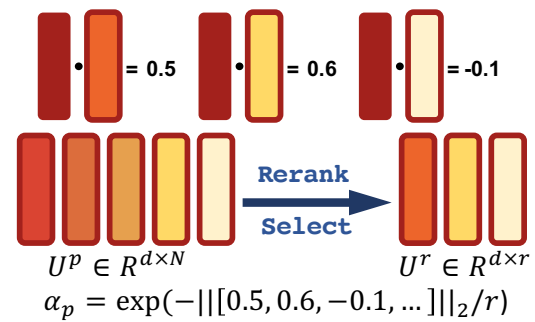
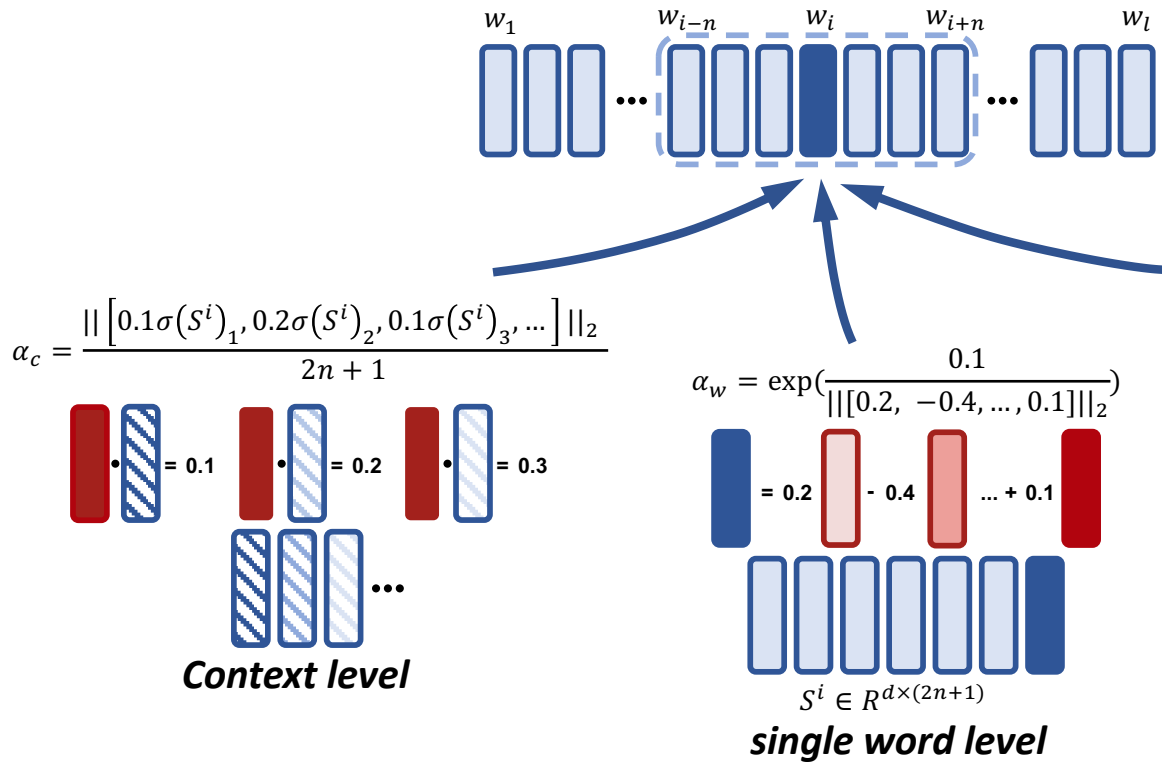


$$\alpha_p = \exp(-\|[0.5, 0.6, -0.1, \dots]\|_2/r)$$

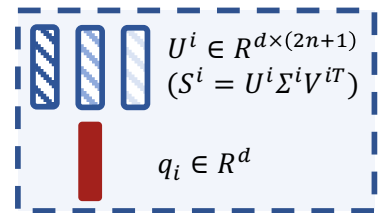
sentence corpus level



Geometric Embedding (GEM)



sentence corpus level



$$v = \sum (\alpha_c + \alpha_w + \alpha_p) w_i$$

Experiments: Unsupervised Tasks

STS sentence similarity datasets

Non-parameterized models	dev	test
GEM + L.F.P (ours)	83.5	78.4
GEM + LexVec (ours)	81.9	76.5
SIF (Arora et al., 2017)	80.1	72.0
uSIF (Ethayarajh, 2018)	84.2	79.5
LexVec	58.78	50.43
L.F.P	62.4	52.0
word2vec skipgram	70.0	56.5
Glove	52.4	40.6
ELMo	64.6	55.9
Parameterized models		
PARAMNT-50M (Wieting and Gimpel, 2017a)	-	79.9
Reddit + SNLI (Yang et al., 2018)	81.4	78.2
GRAN (Wieting and Gimpel, 2017b)	81.8	76.4
InferSent (Conneau et al., 2017)	80.1	75.8
Sent2Vec (Pagliardini et al., 2018)	78.7	75.5
Paragram-Phrase (Wieting et al., 2015a)	73.9	73.2

Table 1: Pearson’s $r \times 100$ on STSB. Best results are in bold.

Experiments: Supervised Tasks

Fix the sentence embeddings and train neural structures for downstream tasks.

Model	Dim	Training time (h)	MR	CR	SUBJ	MPQA	SST	TREC	MRPC	SICK-R	SICK-E
<i>Non-parameterized models</i>											
GEM + L.F.P	900	0	79.8	82.5	93.8	89.9	84.7	91.4	75.4/82.9	86.5	86.2
GEM + GloVe	300	0	78.8	81.1	93.1	89.4	83.6	88.6	73.4/82.3	86.3	85.3
SIF	300	0	77.3	78.6	90.5	87.0	82.2	78.0	-	86.0	84.6
uSIF	300	0	-	-	-	-	80.7	-	-	83.8	81.1
p-mean	3600	0	78.4	80.4	93.1	88.9	83.0	90.6	-	-	-
GloVe BOW	300	0	78.7	78.5	91.6	87.6	79.8	83.6	72.1/80.9	80.0	78.6
<i>Parameterized models</i>											
InferSent	4096	24	81.1	86.3	92.4	90.2	84.6	88.2	76.2/83.1	88.4	86.3
Sent2Vec	700	6.5	75.8	80.3	91.1	85.9	-	86.4	72.5/80.8	-	-
SkipThought-LN	4800	336	79.4	83.1	93.7	89.3	82.9	88.4	-	85.8	79.5
FastSent	300	2	70.8	78.4	88.7	80.6	-	76.8	72.2/80.3	-	-
<i>à la carte</i>	4800	N/A	81.8	84.3	93.8	87.6	86.7	89.0	-	-	-
SDAE	2400	192	74.6	78.0	90.8	86.9	-	78.4	73.7/80.7	-	-
QT	4800	28	82.4	86.0	94.8	90.2	87.6	92.4	76.9/84.0	87.4	-
STN	4096	168	82.5	87.7	94.0	90.9	83.2	93.0	78.6/84.4	88.8	87.8
USE	512	N/A	81.36	86.08	93.66	87.14	86.24	96.60	-	-	-

Summary

- Identify the new semantic meaning q_i by QR factorization
- Measure the importance of a word from 3 levels based on q_i

Q&A

Thank you!