CS224N Research Highlight

A Simple but Tough-to-beat Baseline for **Sentence Embeddings**

Sanjeev Arora, Yingyu Liang, Tengyu Ma
Princeton University
In submission to ICLR 2017

Presenter: Danqi Chen

Word ⟶ Sentence ?

$$\text{linguistics} = \begin{pmatrix} 0.286 \\ 0.792 \\ -0.177 \\ -0.107 \\ 0.109 \\ -0.542 \\ 0.349 \\ 0.271 \end{pmatrix}$$

Word ⟶ Sentence ?

linguistics = $\begin{pmatrix} 0.286 \\ 0.792 \\ -0.177 \\ -0.107 \\ 0.109 \\ -0.542 \\ 0.349 \\ 0.271 \end{pmatrix}$

Natural language processing is fun. = $\begin{pmatrix} -0.132 \\ 1.129 \\ 0.827 \\ 0.110 \\ -0.527 \\ 0.156 \\ 0.349 \\ -0.286 \end{pmatrix}$

Sentence embedding

- Compute **sentence similarity** using the inner product:

S1: Mexico wishes to guarantee citizen's safety.

S2: Mexico wishes to avoid more violence.

Score: 4 (/5)

S1: Iranians Vote in Presidential Election.

S2: Keita Wins Mali Presidential Election.

Score: 0.4 (/5)

Sentence embedding

- Use as features for **sentence classification** (e.g., sentiment analysis):

Natural language processing is fun. $=$ $\begin{pmatrix} -0.132 \\ 1.129 \\ 0.827 \\ 0.110 \\ -0.527 \\ 0.156 \\ 0.349 \\ -0.286 \end{pmatrix}$ → 

From Bag-of-words to Complex Models…

- Bag-of-words (BoW)

v(“natural language processing”) =

1/3 (v(“natural”) + v(“language”) + v(“processing”))

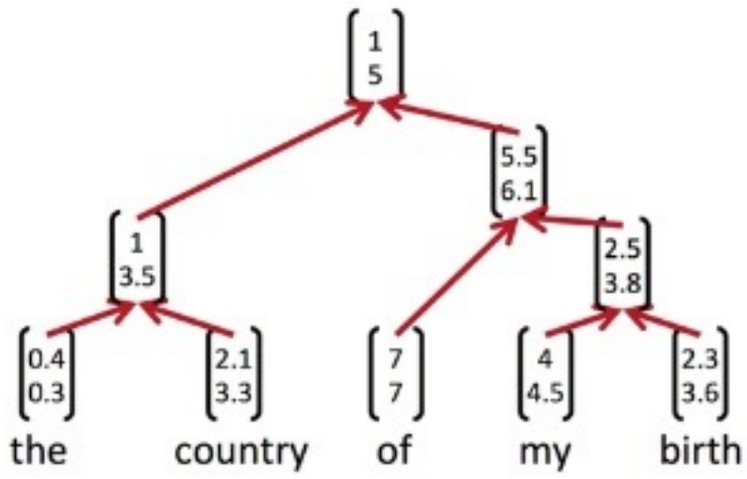
From Bag-of-words to Complex Models…

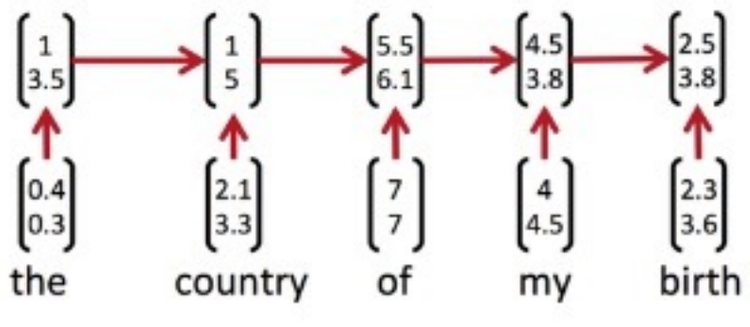
- Bag-of-words (BoW)

v(“natural language processing”) =

1/3 (v(“natural”) + v(“language”) + v(“processing”))

- Recurrent neural networks, recursive neural networks, convolutional neural networks..





This paper

- A VERY SIMPLE unsupervised method
  - **weighted Bag-of-words + remove some special direction**

This paper

- A VERY SIMPLE unsupervised method
  - **weighted Bag-of-words + remove some special direction**
- **Step 1**:

for all sentence s in $\mathcal{S}$ **do**

$$v_s \leftarrow \frac{1}{|s|} \sum_{w \in s} \frac{a}{a+p(w)} v_w$$

end for

This paper

- A VERY SIMPLE unsupervised method
  - **weighted Bag-of-words + remove some special direction**
- **Step 1**:

for all sentence s in $\mathcal{S}$ **do**

$$v_s \leftarrow \frac{1}{|s|} \sum_{w \in s} \frac{a}{a+p(w)} v_w$$

end for

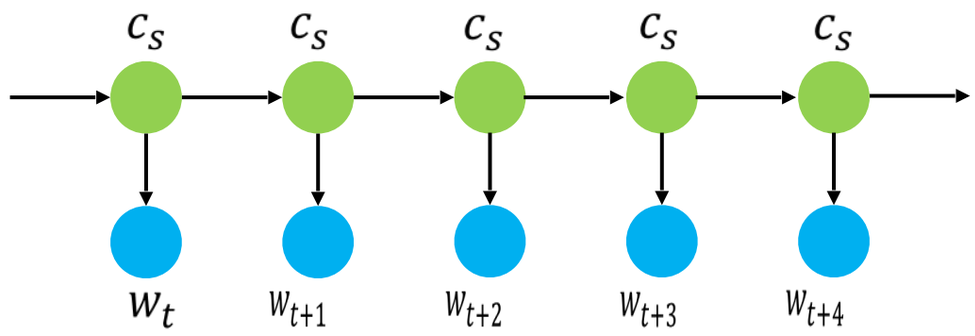
- **Step 2**:

Compute the first principal component u of $\{v_s : s \in \mathcal{S}\}$
for all sentence s in $\mathcal{S}$ **do**

$$v_s \leftarrow v_s - uu^\top v_s$$

end for

A Probabilistic Interpretation



$$\Pr[w_t \mid c_s] = \alpha p(w_t) + (1-\alpha)\frac{\exp\langle v_{w_t}, b_s\rangle}{Z_{\tilde{c}_s}}$$

Smoothing term:
w_t is emitted from background probability (irrelevant to the vector c_s)

w_t is emitted according to correlation with the shifted context vector

$$b_s = \beta c_0 + (1-\beta)c_s$$

common discourse, often related to syntax

Results

sentence similarity

	Results collected from (Wieting et al., 2016) except tfidf-GloVe											Our approach	
Supervised or not	Su.							Un.			Se.	Un.	Se.
Tasks	PP	PP -proj.	DAN	RNN	iRNN	LSTM (no)	LSTM (o.g.)	ST	avg-GloVe	tfidf-GloVe	avg-PSL	GloVe +WR	PSL +WR
STS'12	58.7	**60.0**	56.0	48.1	58.4	51.0	46.4	30.8	52.5	58.7	52.8	56.2	59.5
STS'13	55.8	56.8	54.2	44.7	56.7	45.2	41.5	24.8	42.3	52.1	46.4	56.6	**61.8**
STS'14	70.9	71.3	69.5	57.7	70.9	59.8	51.5	31.4	54.2	63.8	59.5	68.5	**73.5**
STS'15	75.8	74.8	72.7	57.2	75.6	63.9	56.0	31.0	52.7	60.6	60.0	71.7	**76.3**
SICK'14	71.6	71.6	70.7	61.2	71.2	63.9	59.0	49.8	65.9	69.4	66.4	72.2	**72.9**
Twitter'15	52.9	52.8	**53.7**	45.1	52.9	47.6	36.1	24.7	30.3	33.8	36.3	48.0	49.0

sentence classification

	PP	DAN	RNN	LSTM (no)	LSTM (o.g.)	skip-thought	Ours
similarity (SICK)	84.9	85.96	73.13	85.45	83.41	85.8	**86.03**
entailment (SICK)	83.1	84.5	76.4	83.2	82.0	-	**84.6**
sentiment (SST)	79.4	83.4	86.5	86.6	**89.2**	-	82.2

Results

sentence similarity

	Results collected from (Wieting et al., 2016) except tfidf-GloVe											Our approach	
Supervised or not	Su.							Un.			Se.	Un.	Se.
Tasks	PP	PP -proj.	DAN	RNN	iRNN	LSTM (no)	LSTM (o.g.)	ST	avg-GloVe	tfidf-GloVe	avg-PSL	GloVe +WR	PSL +WR
STS'12	58.7	**60.0**	56.0	48.1	58.4	51.0	46.4	30.8	52.5	58.7	52.8	56.2	59.5
STS'13	55.8	56.8	54.2	44.7	56.7	45.2	41.5	24.8	42.3	52.1	46.4	56.6	**61.8**
STS'14	70.9	71.3	69.5	57.7	70.9	59.8	51.5	31.4	54.2	63.8	59.5	68.5	**73.5**
STS'15	75.8	74.8	72.7	57.2	75.6	63.9	56.0	31.0	52.7	60.6	60.0	71.7	**76.3**
SICK'14	71.6	71.6	70.7	61.2	71.2	63.9	59.0	49.8	65.9	69.4	66.4	72.2	**72.9**
Twitter'15	52.9	52.8	**53.7**	45.1	52.9	47.6	36.1	24.7	30.3	33.8	36.3	48.0	49.0

sentence classification

	PP	DAN	RNN	LSTM (no)	LSTM (o.g.)	skip-thought	Ours
similarity (SICK)	84.9	85.96	73.13	85.45	83.41	85.8	**86.03**
entailment (SICK)	83.1	84.5	76.4	83.2	82.0	-	**84.6**
sentiment (SST)	79.4	83.4	86.5	86.6	**89.2**	-	82.2

Thanks!