Learning Sentence Representation with Guidance of Human Attention

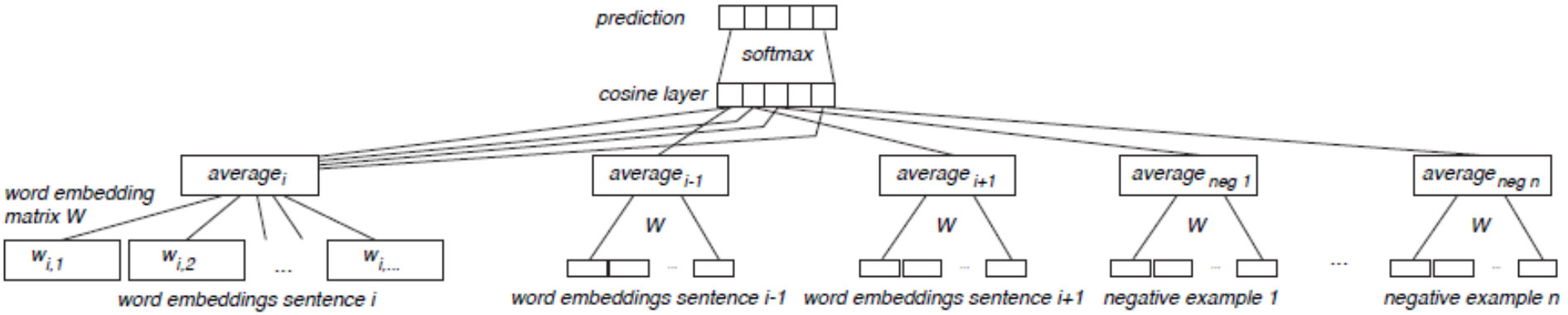
2017-12-06

Background

- Human read sentences by fixation and saccades
- Based on Siamese CBOW
- Assign different weight to the vectors

Siamese CBOW

- Based on CBOW
- Averaging the embedding of words in a sentence
- Designed specifically for the task of averaging them.



Siamese Network modal

- Training objective

$$p_\theta(s_i, s_j) = \frac{e^{\cos(\mathbf{s_i^\theta}, \mathbf{s_j^\theta})}}{\sum_{s_\prime \in S} e^{\cos(\mathbf{s_i^\theta}, \mathbf{s_\prime^\theta})}} \qquad p(s_i, s_j) = \begin{cases} \frac{1}{|S^+|}, & \text{if } s_j \in S^+ \\ 0, & \text{if } s_j \in S^-. \end{cases}$$

$$L = -\sum_{s_j \in \{S^+ \cup S^-\}} p(s_i, s_j) \cdot \log(p_\theta(s_i, s_j))$$

$$g_{sentence}(x) = \frac{1}{n}\sum_{i=1}^{n} W_w^{x_i}, \qquad g_{sentence}(x) = \frac{1}{n}\sum_{i=1}^{n} attention(x_i) W_w^{x_i}$$

Human Attention

- Surprisal
- self-information, measures the amount of information conveyed by the target

$$s^{x_t} = -log(P(x_t|x_1, ..., x_{t-1})),$$

$$attention(x_t) = \frac{exp(s^{x_t})}{\sum_{i \in [1,...,n]} exp(s^{x_i})},$$

Human Attention

- POS tags
- give more weights to words with NN and VBG tags, and less weights to words with DT and IN tags.

$$attention(x_t) = \frac{exp(W_w^{x_t} \cdot W_c^{x_t})}{\sum_{i \in [1,..,n]} exp(W_w^{x_t} \cdot W_c^{x_t})},$$

Evaluation

	SCBOW				
	Base	TF-IDF	ATT-SUR	ATT-POS	ATT-CCG
MSRpar	0.429	0.412	**0.437**	0.414	0.419
MSRvid	0.620	0.611	0.672	0.702	**0.734**
OnWN	0.687	0.688	0.677	0.695	**0.696**
SMTeurop	0.537	0.542	0.533	0.538	**0.552**
SMTnews	0.523	0.525	0.544	0.541	**0.557**
2012 Average	0.559	0.556	0.573	0.578	**0.592**
FNWN	0.378	0.375	0.350	0.383	**0.392**
OnWN	0.584	0.585	**0.649**	0.609	0.583
headlines	0.693	0.688	0.705	0.704	**0.711**
2013 Average	0.552	0.549	**0.568**	0.565	0.562

Dataset： SemEval

data sample: He is smart = He is a wise man.

Microsoft to acquire Linkedin ≠ Linkedin to acquire microsoft

Evaluation: Pearson's r (Spearman's r)