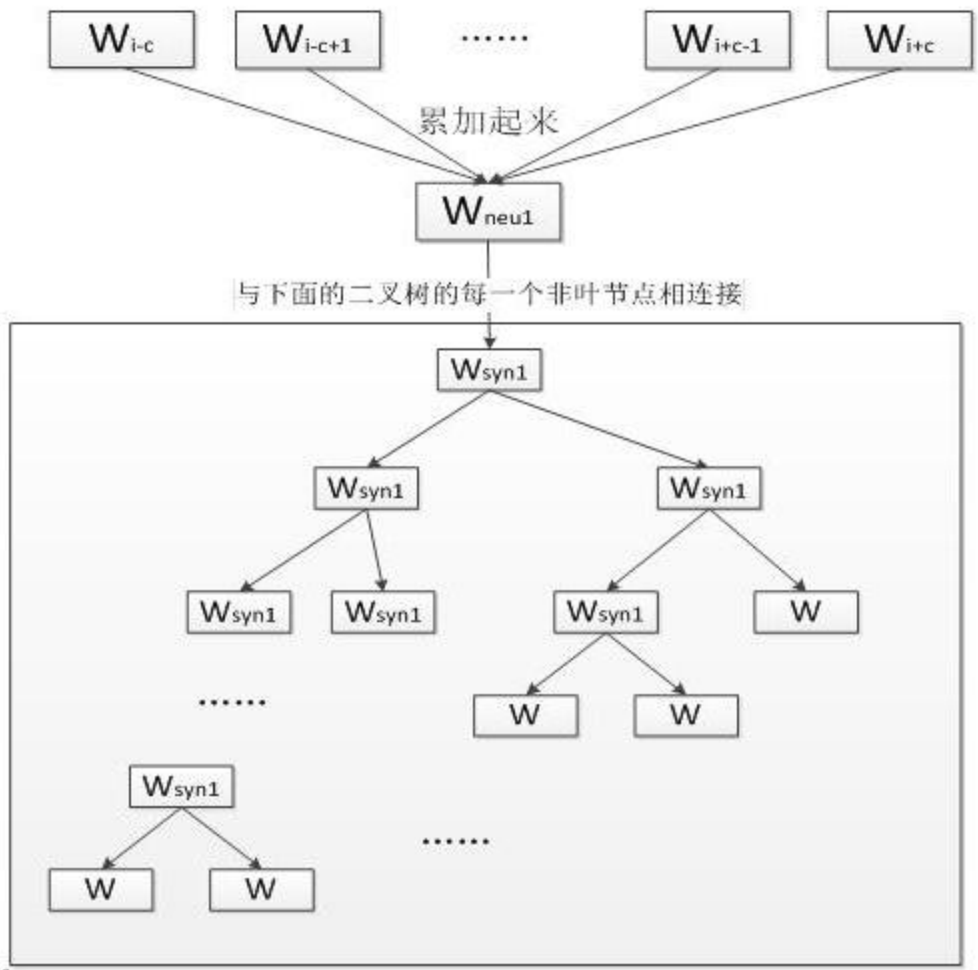
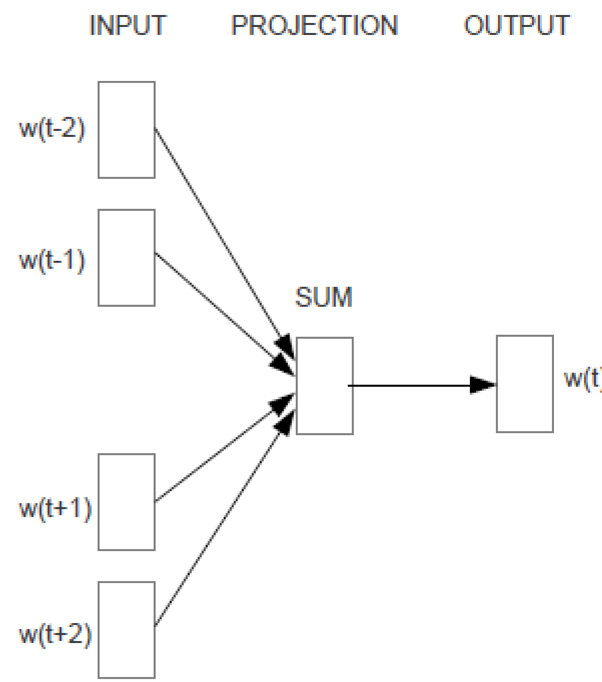
Siamese CBOW : Optimizing Word Embeddings for Sentence Representation

2017-11-29

Background

- Based on CBOW

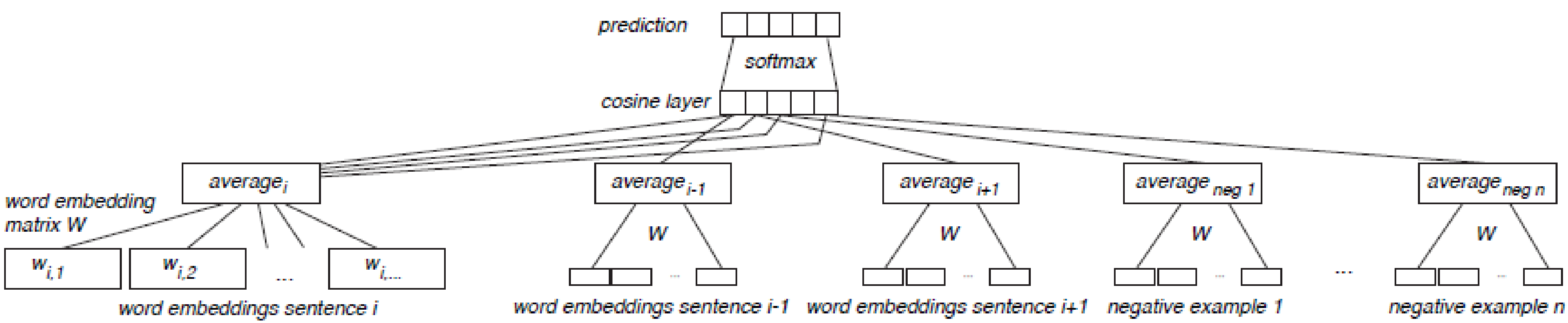




CBOW

Siamese CBOW

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



Network modal

- Training objective

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

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

- a high cosine similarity to positive examples
- a low cosine similarity to negative examples.

Evaluation

Dataset	w2v skipgram	w2v CBOW	skip-thought	Siamese CBOW
2012				
MSRpar	.3740 (.3991)	.3419 (.3521)	.0560 (.0843)	**.4379**† (.4311)
MSRvid	.5213 (.5519)	.5099 (.5450)	**.5807** (.5829)	.4522† (.4759)
OnWN	.6040 (.6476)	.6320 (.6440)	.6045 (.6431)	**.6444**† (.6475)
SMTeuroparl	.3071 (.5238)	.3976 (.5310)	.4203 (.4999)	**.4503**† (.5449)
SMTnews	**.4487** (.3617)	.4462 (.3901)	.3911 (.3628)	.3902† (.4153)

Skip-thought: learning sentence representation by using RNN, taking word order into account.

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)

Conclusion

	CBOW	Siamese CBOW
Input	Random vector	Word2vec word embedding
production	Word embedding	Word embedding
Design	Predict a word given context	Predict a sentence given positive sample and negative sample
application	General	Specifically for averaging word embedding