Learning Neural Word Salience Scores

Krasen Samardzhiev, Andrew Gargett, Danushka Bollegala

University of Liverpool

Problem

Humans can easily recognise the words that contribute to the meaning of a sentence (i.e. content words) from words that serve only a grammatical functionality (i.e. functional words). If we can accurately compute the *salience* of words, then we can develop better representations of texts that can be used in downstream NLP tasks such as similarity measurement or text (e.g. sentiment, entailment) classification.

Contributions

We define the *salience* $q(w)$ of a word w in a given text T as the semantic contribution made by w towards the overall meaning of T. We propose a method that first randomly initialises these scores, and subsequently updates them such that we can accurately predict the words in local contexts. The method:

- Learns a salience score for each word from a corpus, and is not based on heuristics.
- It does *not* require labelled data for learning NWS scores. It only needs semantically similar and semantically dissimilar pairs of sentences for learning the NWS scores, which are automatically extracted from the given training corpus.
- Requires significantly less training and prediction times than similar methods that aim to learn sentence embeddings.

References

References

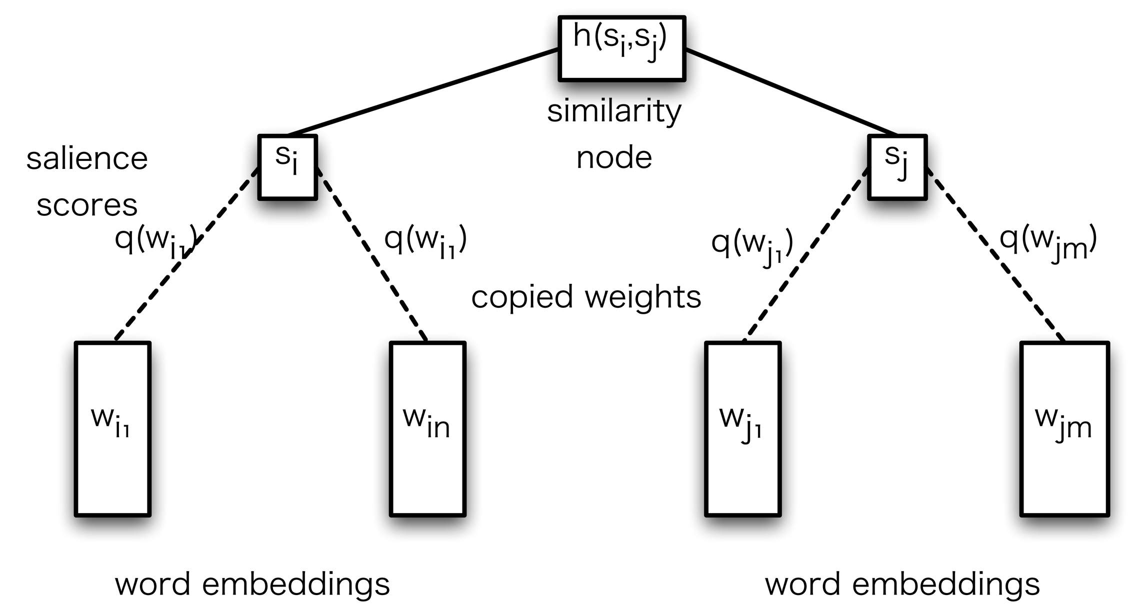
[1] S. Arora, Y. Liang, and T. Ma. A simple but tough-to-beat baseline for sentence embeddings. In *Proc. of International Conference on Learning Representations (ICLR)*, 2017.

[2] M. U. Gutmann and A. Hyvärinen. Noise-contrastive estimation of unnormalized statistical models, with applications to natural image statistics. *Journal of Machine Learning Research*, 13:307 – 361, February 2012.

[3] T. Kenter, A. Borisov, and M. de Rijke. Siamese cbow: Optimizing word embeddings for sentence representations. In *Proc. of the Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 941–951, 2016.

[4] R. Kiros, Y. Zhu, R. Salakhutdinov, R. S. Zemel, A. Torralba, R. Urtasun, and S. Fidler. Skip-thought vectors. In *Proc. of Advances in Neural Information Processing Systems (NIPS)*, pages 3276–3284, 2015.

Model



We take pre-trained word embeddings of the words in a sentence S_i as the input and compute a weighted average $\boldsymbol{S}_i$, where the weights correspond to the word salience scores of the words in S_i. We expect adjacent sentences to be topically related and so their vector representations should have a high degree of similarity.
The similarity between two sentence embeddings is computed using a softmax function:

$$h(\boldsymbol{s}_i, \boldsymbol{s}_j) = \frac{e^{\cos(\boldsymbol{s}_i, \boldsymbol{s}_j)}}{\sum_{S_k \in C} e^{\cos(\boldsymbol{s}_i, \boldsymbol{s}_k)}}$$

Following noise-contrastive estimation[2], we approximate the normalization term. The model is trained using the two adjacent sentences to $\mathcal{S}_i$ as positive examples, and two negative examples sampled uniformly from the whole corpus .The loss function is defined as:

$$L = t\log\left(h(\boldsymbol{s}_i, \boldsymbol{s}_j)\right) + (1-t)\log\left(1 - h(\boldsymbol{s}_i, \boldsymbol{s}_j)\right)$$

$$t = \begin{cases} 1 & j = i+1 \\ 1 & j = i-1 \\ 0 & otherwise \end{cases}$$

Sample Words

Table 1: Words with low salience scores

ISF	NWS (ISF init.)	NWS (rand init.)
the	your	alexis
to	our	tobias
i	we	copyright
and	my	rupert
a	you	spotted
of	us	vehicle
was	me	sword
he	i	isaac
his	voice	fletcher
you	has	cook

Table 2: Words with high salience scores

ISF	NWS (ISF init.)	NWS (rand init.)
pathways	guess	hurdling
conspiratorial	boulder	happen
henna	autopsy	weird
alejandro	hippy	alejo
bedpost	alejandro	bolivians
swiveling	philosophy	his
confederate	arrow	answer
mid-morning	germany	her
alejo	spotted	yesterday
phd	bookstore	replied

Table 3: Correlations with psycholingusitic scores

Embed.	Arousal	Conc.	Dom.	Img.	Valance
GloVe	0.03	0.26	0.09	0.25	0.03
CBOW	0.04	-0.35	-0.04	-0.37	0.04
SGNS	-0.01	0.27	0.06	0.27	-0.01

Performance in sentence similiarity tasks

Dataset	SMOOTH[1]	skip-thought[4]	Siamese-CBOW[3]	AVG	ISF	NWS
2012						
MSRpar	43.6	5.6	**43.8**	28.4	39.1	28.5
OnWN	54.3	60.5	64.4	47.1	60.5	**65.5***
SMTeuroparl	**51.1***	42.0	45.0	37.1	44.5	50.1
SMTnews	42.2	39.1	39.0	32.2	34.9	**44.7***
2013						
FNWN	23.0	**31.2**	23.2	26.9	29.4	25.2
OnWN	68.0*	24.2	49.9	25.0	63.2	**78.1***
headlines	63.8	38.6	**65.3***	40.2	59.4	57.0
2014						
OnWN	68.0	46.8	60.7	41.1	68.5	**80.8***
deft-forum	29.1	37.4	**40.8**	27.1	37.1	29.9
deft-news	**68.5**	46.2	59.1	48.8	63.6	65.4
headlines	59.3	40.3	**63.6***	41.9	58.8	56.2
images	74.1*	42.6	65.0	35.3	66.3	**75.9***
tweet-news	57.3	51.4	**73.2***	41.7	57.1	64.5*
2015						
answers-forums	41.4	27.8	21.8	25.7	37.6	**49.6***
answers-students	61.5	26.6	36.7	56.5	67.1	**68.0**
belief	47.7	45.8	47.7	29.3	43.2	**54.3***
headlines	64.0	12.5	21.5	49.3	**65.4**	65.3
images	75.4*	21	25.6	49.8	66.1	**76.6***
Overall Average	55.1	35.5	47.0	38.0	53.4	**57.6**