Molding CNNs for text: non-linear, non-consecutive convolutions

Tao Lei, Regina Barzilay and Tommi Jaakkola
MIT, CSAIL
{taolei, regina, tommi}@csail.mit.edu

Motivation

The success of deep learning often derives from well-chosen operational building blocks.

Question: can we design neural network components better for text processing?

This work

Motivated by previous NLP methods like string kernels, we revise the feature mapping operation (i.e. convolution operation) of CNNs

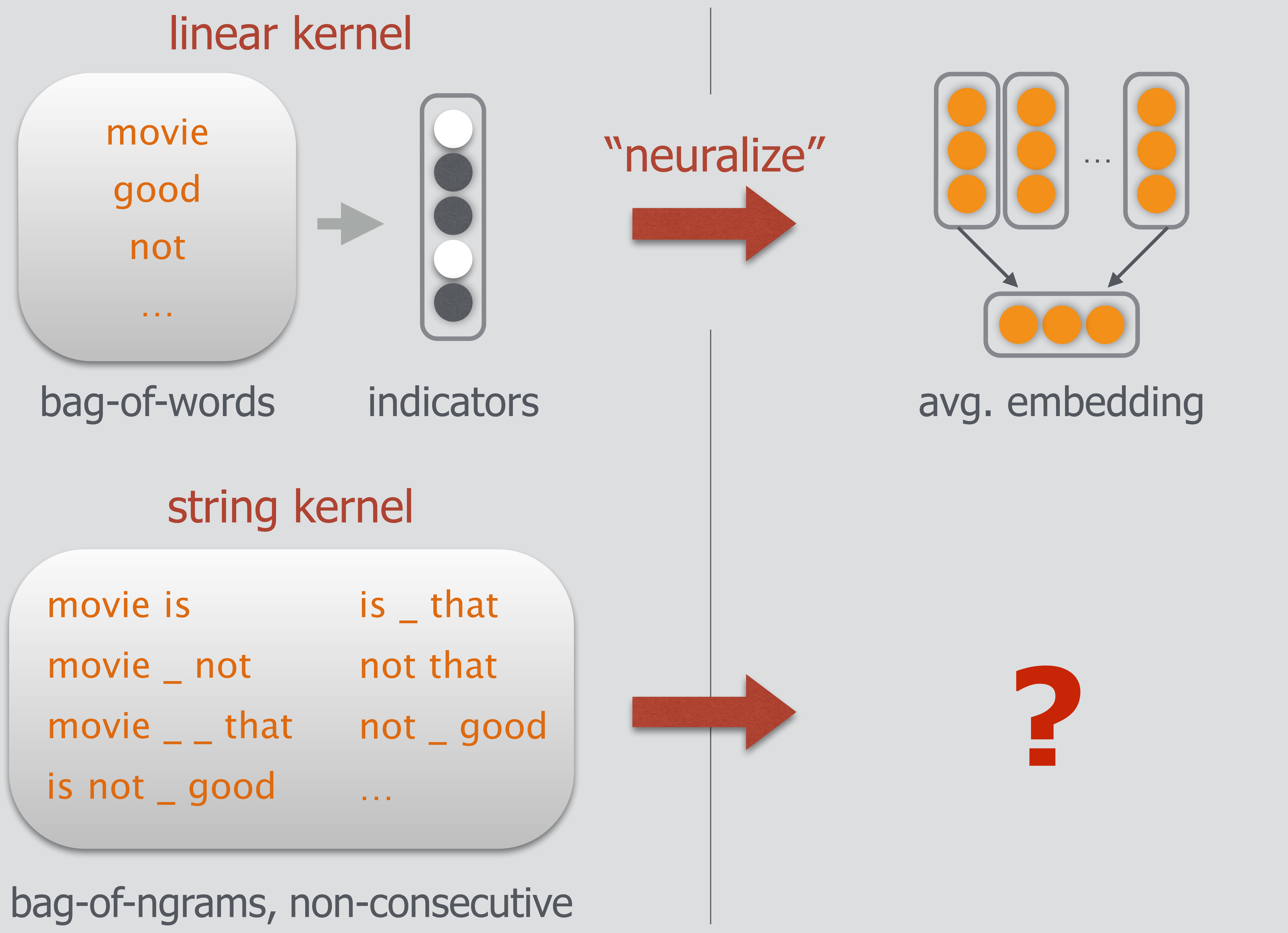
- Directly handles non-consecutive n-gram patterns, e.g. "not nearly as good" etc.
- Use tensor algebra to capture n-gram interactions

Our code and data are available at
https://github.com/taolei87/text_convnet

Example

Consider generating the feature representation of the following sentence:

"the movie is not that good"

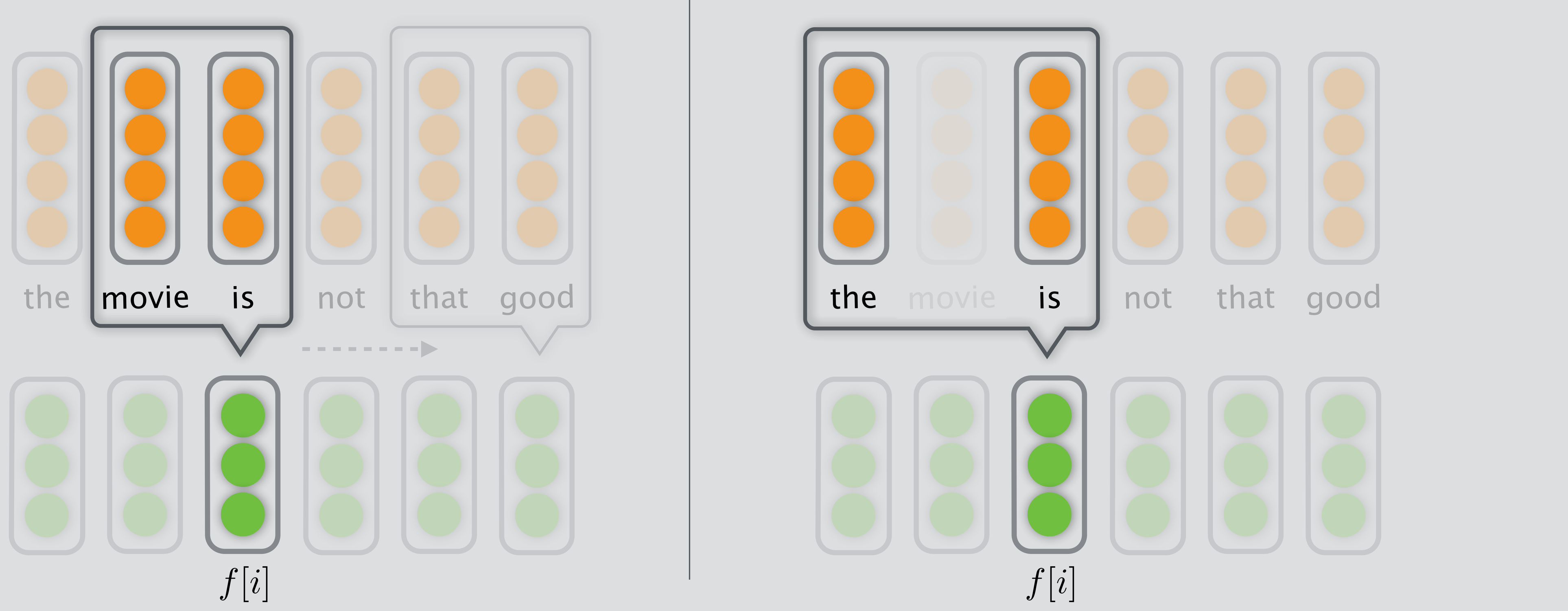


Model

Apply the "string kernel" idea to CNN feature mapping. 2gram case:

(i) non-linear high-order filters

(ii) averaging non-consecutive ngrams



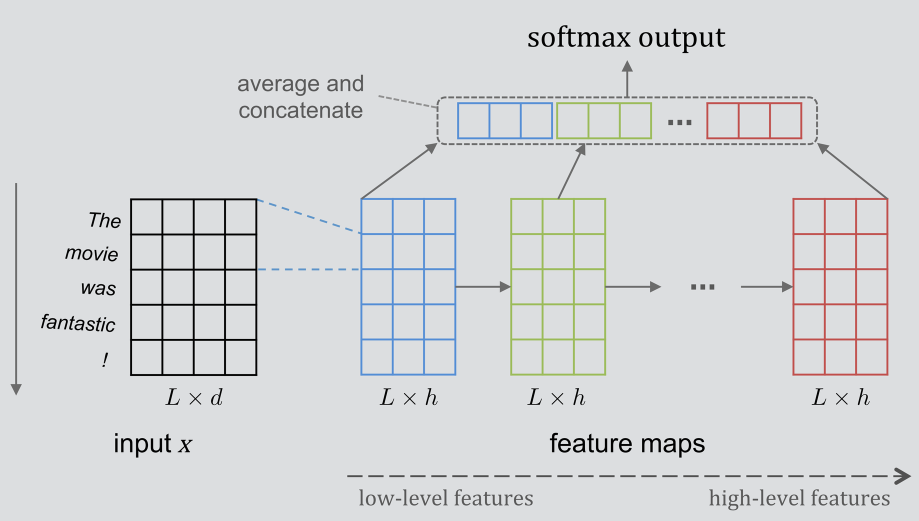
$$f[i] \leftarrow \sum_{j<i} \lambda^{i-j-1}\, \mathbf{T} \cdot (\mathbf{x}_j \otimes \mathbf{x}_i) \quad \text{(2-gram)}$$

$$f[i] \leftarrow \sum_{k<j<i} \lambda^{i-k-2}\, \mathbf{T} \cdot (\mathbf{x}_k \otimes \mathbf{x}_j \otimes \mathbf{x}_i) \quad \text{(3-gram)}$$

(iii) linear time dynamic programming possible when **T** is low-rank factorized!

Architecture

- Directly plug into CNNs for feature extraction
- Can be stacked or feed into activation cells



Results

Model evaluated on sentiment analysis task, newswire and POS classification tasks.

Model	Fine	Binary	Time
DCNN [1]	48.5	86.9	-
DNN-MC [2]	47.4	88.1	156
RLSTM [3]	51.0	88.0	164
Ours (best)	52.7	88.6	28
(avg.)	51.4	88.4	

Table 1: Results on Stanford Sentiment Treebank.

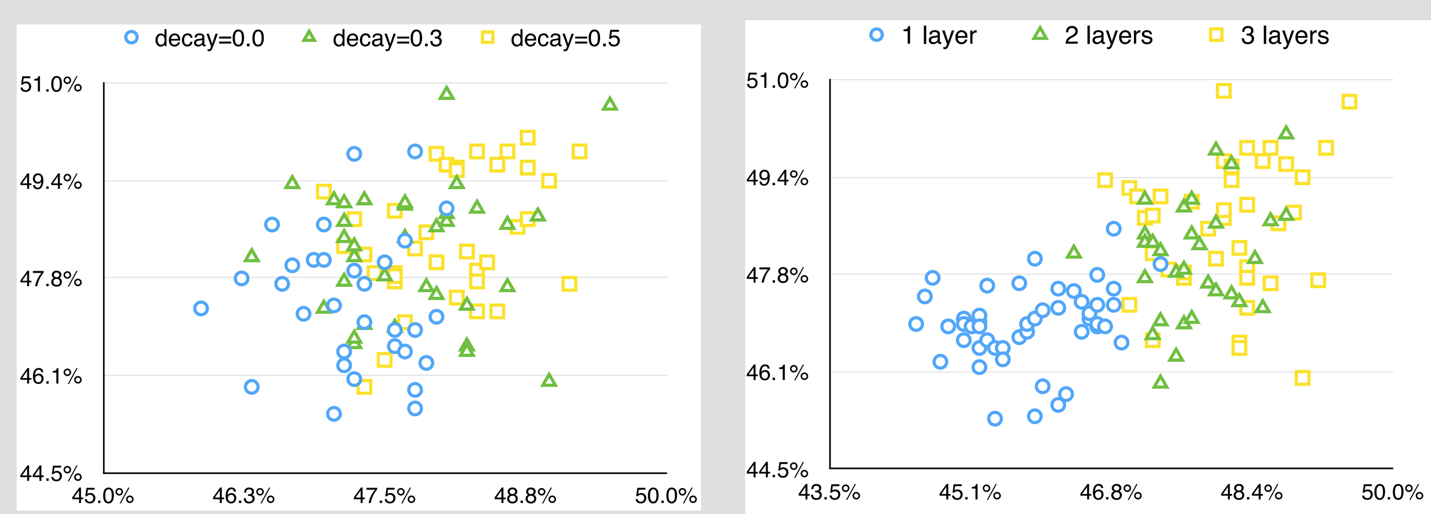


Figure 1: Analysis of our model. (a) better acc% when handles non-consecutive ngrams; (b) deeper model gives better acc%.

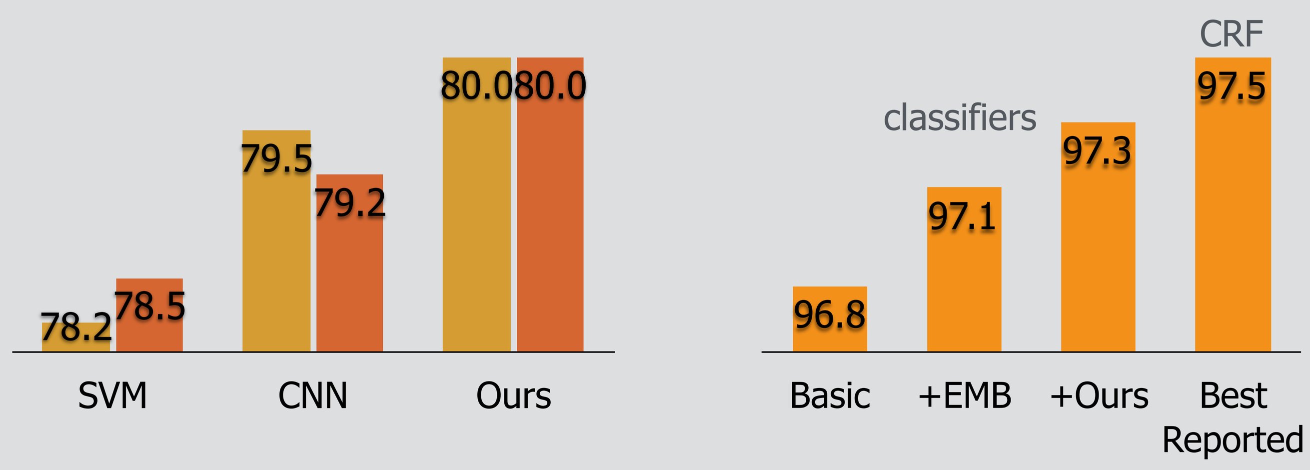


Figure 2: Results on document classification (left) and POS classification (right).

References

[1] Nal Kalchbrenner, Edward Grefenstette, and Phil Blunsom. A convolutional neural network for modelling sentences. ACL 2014

[2] Yoon Kim. Convolutional neural networks for sentence classification. EMNLP 2014

[3] Kai Sheng Tai, Richard Socher, and Christopher D Manning. Improved semantic representations from tree-structured long short-term memory networks. ACL 2015