CS224n: NLP with Deep Learning **Winter 2019**

Lecture 11

Course Coordinator: Prof. Chris Manning *Scribes: Akash Gupta*

Problems with RNNs -

- RNNs cannot capture phrases without prefixes
- Often capture too much of last words in final vector.

1D convolution for text -

- Represent each word with a fixed dimension vector.
- Apply several filter to that matrix
- Apply techniques like padding, stride, dilated convolution, k-max pooling
- Max pooling - summarizes the idea in a particular part of the sentence.

CNNs for Sentence Classification - *Yoon Kim 2014. ConvNets for Sentence Classification*

- Start by word vectors of length k.
- Sentence is made by concatenating the word vectors.
- Convolution filter is represented as 1D vector.
- Use filters of different sizes.

− > CNNs are easy to parallelize on GPUs allowing faster computation whereas RNNs are slow.

− > CNN Translation - *Kalchbrenner and Blunsom 2013. Recurrent Continuous Translation models.* First modern Machine Translation paper. CNNs for encoder and sequence model in the decoder.

− > *Conneau et al. 2017. Very Deep CNNs for Text Classification.* Built very deep networks for NLP (Like Resnets in Vision). Works over characters.

− >*Bradbury et al. 2017. Quasi-RNN* - gives psuedo-recurrence ability. Convolutions are done using previous and current time step. Faster than RNNs but suffer with performance.