BUS 243

Lecture 9: CNN and review

The non-linearity of the network, as well as the ability to easily integrate pre-trained word embeddings, often lead to superior classification accuracy – Yoav Goldberg

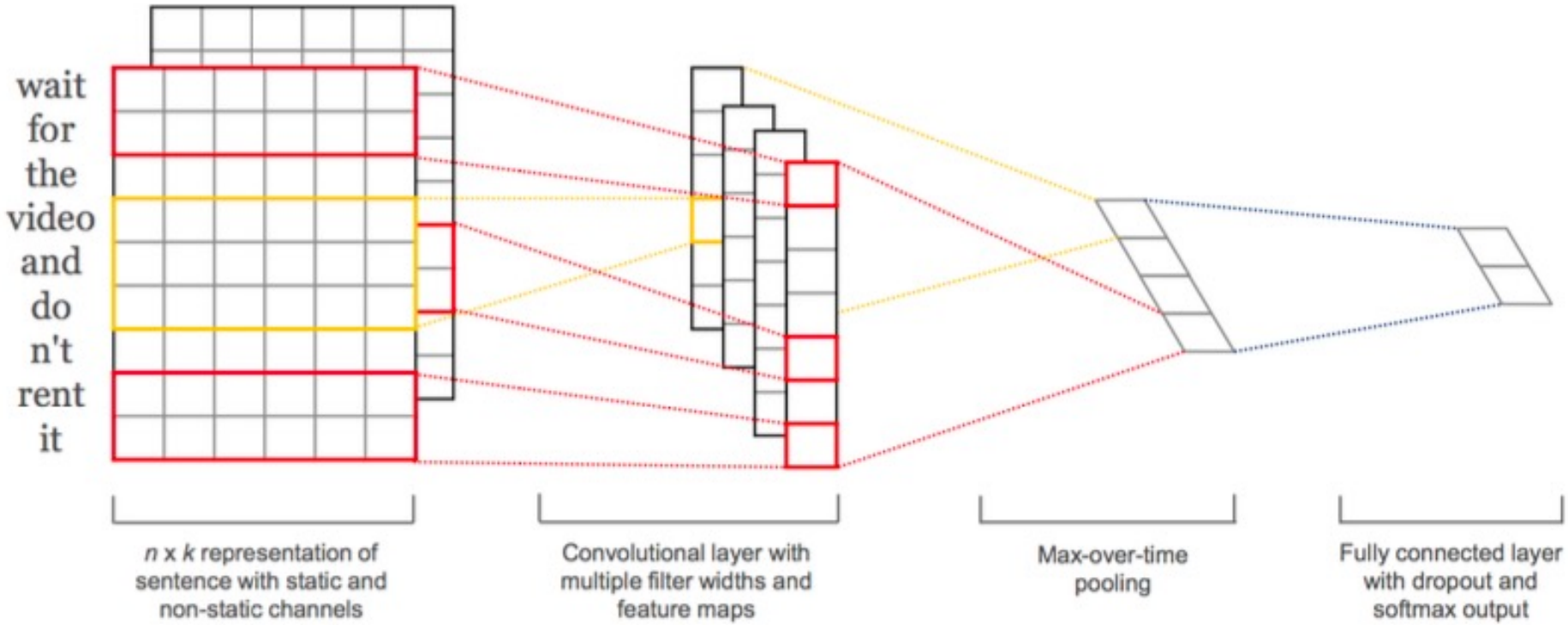
CONVOLUTION NEURAL NETS

- However, you will see that the model with pre-trained embedding in homework 5 perform worse
- It is not because the embedding, but because the modeling choice
- So far, the NNs considered have all been fully connected
  - What's the meaning?

- Consider image classification task
  - Fully connected model
    - Absolute value of light intensities at global level
  - Hardly make sense in general
    - Differences in pixel values at local level
- CNN is taking account kind of window!

- *Despite little tuning of hyperparameters, a simple CNN with one layer of convolution performs remarkably well. Our results add to the well-established evidence that unsupervised pre-training of word vectors is an important ingredient in deep learning for NLP – Yoon Kim*

- You can get good results for document classification with a single layer CNN
- In a convolutional layer, you can specify the size of the window, often referred to as the *kernel size* or *filter size*
  - small matrix that slides over the input data
  - determines the receptive field or the area of input that the kernel considers at each step


wait
for
the
video
and
do
n't
rent
it
n x k representation of sentence with static and non-static channels
Convolutional layer with multiple filter widths and feature maps
Max-over-time pooling
Fully connected layer with dropout and softmax output

- Some hyperparameters matter more than others
- Unfortunately, a downside is that they require practitioners to specify the exact model architecture to be used and to set the accompanying hyperparameters.
  - Same as demining the size of window

- Using IMDB movie review data, let's see all the necessary steps for CNN classification
- Good for review what we have learned and check where CNN could be used

CNN (OR GENERAL) PIPELINE

- You will see that major part of text classification is on data preparation thanks to deep learning platform
- You have to think 3 steps at least
  - Validation method
    - how to split data?
  - Normalization
    - Data!
  - Vocabulary