1、 Recursive NN require a parser to get tree structure.

2、 Recurrent NN cannot capture phrases without prefix context and often capture too much of last words in final vector.

问题：

（1）在两种 RNN 中是不是每一个词向量就是神经网络的输入，语法分析树就是所需训练的网络。

（2）Recurrent NN 为什么会出现上述情况，cannot capture phrases without prefix context.

RNN：需要依靠语法分析。

CNN：1、计算每一个可能的短语 2、不需要语法分析

Single Layer CNN

1、 convolutional filter: w (goes over h words)

2、 filter w applied to all possible windows 这是 CNN 不需要语法分析的原因。

Single Layer CNN: pooling layer

上一步中计算出来的特征太多，和句子长度一样，所以需要简化，就需要 pooling layer。

In particular： max-over-time pooling layer

Idea: capture most important activation (maximum over time)

按以上的方法做存在一个问题，最后只能得到一个特征，但是我们需要更多的特征。

解决方法：multiple filters

Use multiple filter weight w

Have different window sizes h

Multiple channel idea 没看懂

Start with two copies

Backprop into only set, keep other static

Both channels are added to ci before max-pooling

问题：

（1）Filter 就相当于 RNN 中的权值矩阵？filter w 如何确定，是随机初始值还是经验确定？

（2）Max-over-time pooling layer 是什么意思

(3) maximum over time 是什么意思。

(4)max-pooling 什么意思

Classification after one CNN layer

First one convolution, followed by a max-pooling

To obtain final feature vector

Final softmax layer

Tricks to make it work better: Dropout

Randomly dropout 0 to some of the feature weights z

Create masking vector r of Bernoulli random variables with probability p (a hyperparameter) of being 1

什么是 masking vector