卷积神经网络

CS2916 大语言模型

饮水思源 爱国荣校

https://plms.ai/teaching/index.html

卷积的定义

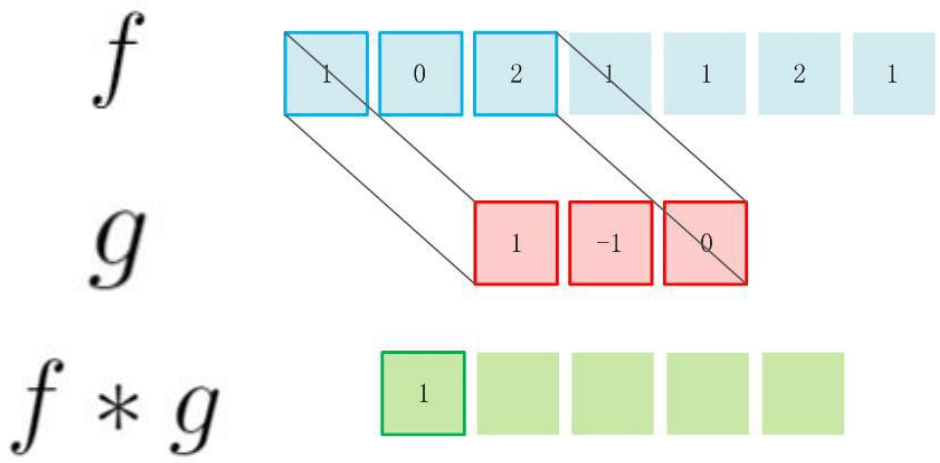
- 数学定义
  - 连续 $(f*g)(t)=\int_{-\infty}^{\infty}f(t-\tau)g(\tau)d\tau$
  - 离散 $(f*g)[n]=\sum_{n=-M}^{M}f[n-m]g[m]$

卷积的定义

- 数学定义
  - 连续 $(f*g)(t)=\int_{-\infty}^{\infty} f(t-\tau)g(\tau)d\tau$
  - 离散 $(f*g)[n]=\sum_{n=-M}^{M} f[n-m]g[m]$

直观的解释

- 数学定义
  - 连续 $(f*g)(t)=\int_{-\infty}^{\infty} f(t-\tau)g(\tau)d\tau$
  - 离散 $(f*g)[n]=\sum_{n=-M}^{M} f[n-m]g[m]$



Input: feature vector

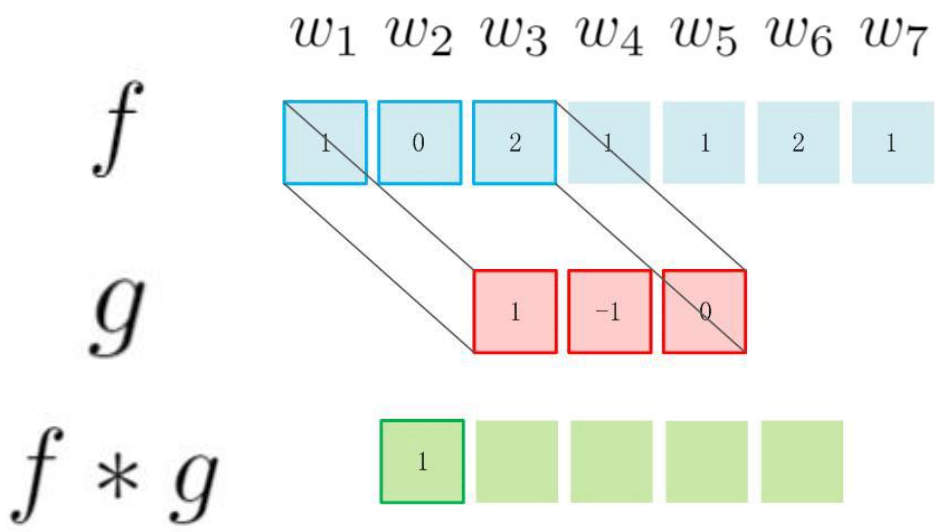
卷积核

Filter: learnable Parameter

Output: hidden vector

CNN中的“先验”

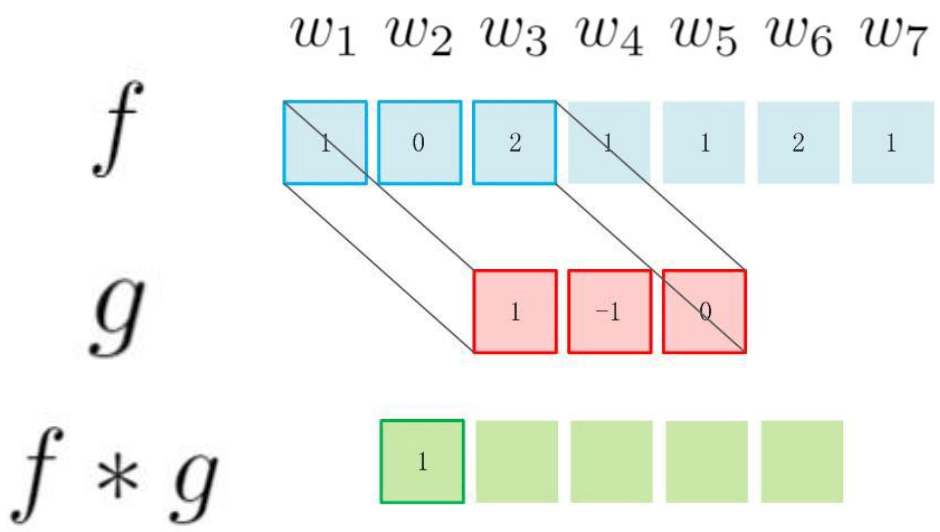
$$(f * g)[n] = \sum_{m=-M}^{M} f[n]g[m]$$



局部假设:不同的单词可以与相邻的单词相互作用

CNN中的“先验”

$$(f * g)[n] = \sum_{m=-M}^{M} f[n]g[m]$$



参数共享:不同词组共享参数

二维卷积

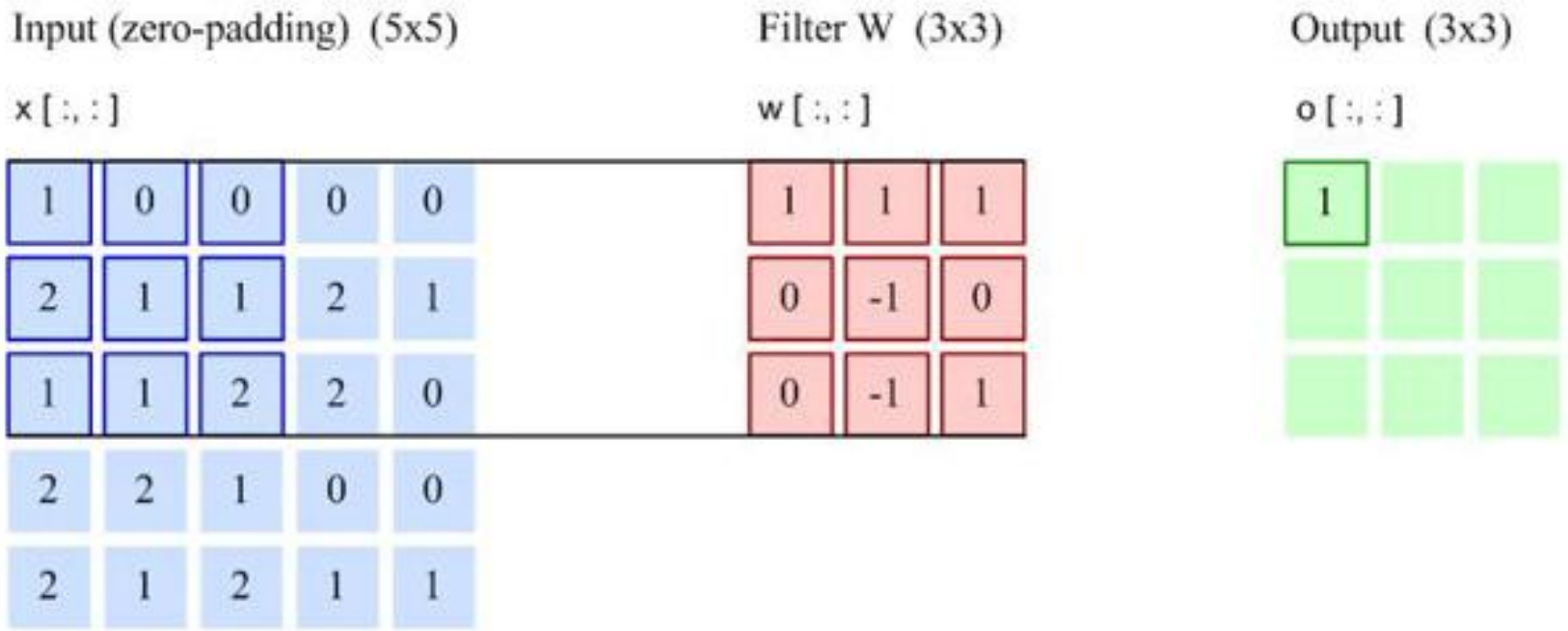
- 处理二维信号、例如图片

$$(f * g)[n] = \sum_{m=-M}^{M} f[n]g[m]$$

二维卷积

- 处理二维信号、例如图片

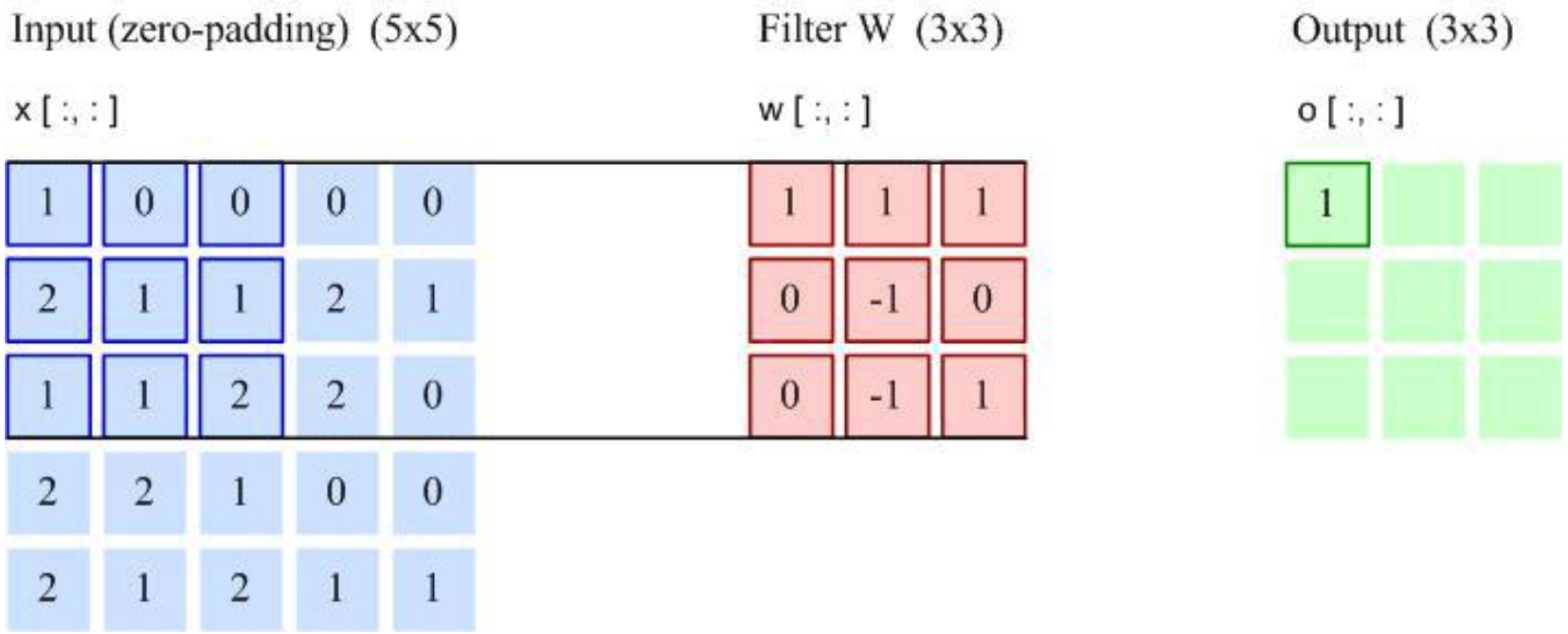
$$(f * g)[n] = \sum_{m=-M}^{M} f[n]g[m]$$



二维卷积

- 处理二维信号、例如图片

$$(f * g)[n] = \sum_{m=-M}^{M} f[n]g[m]$$

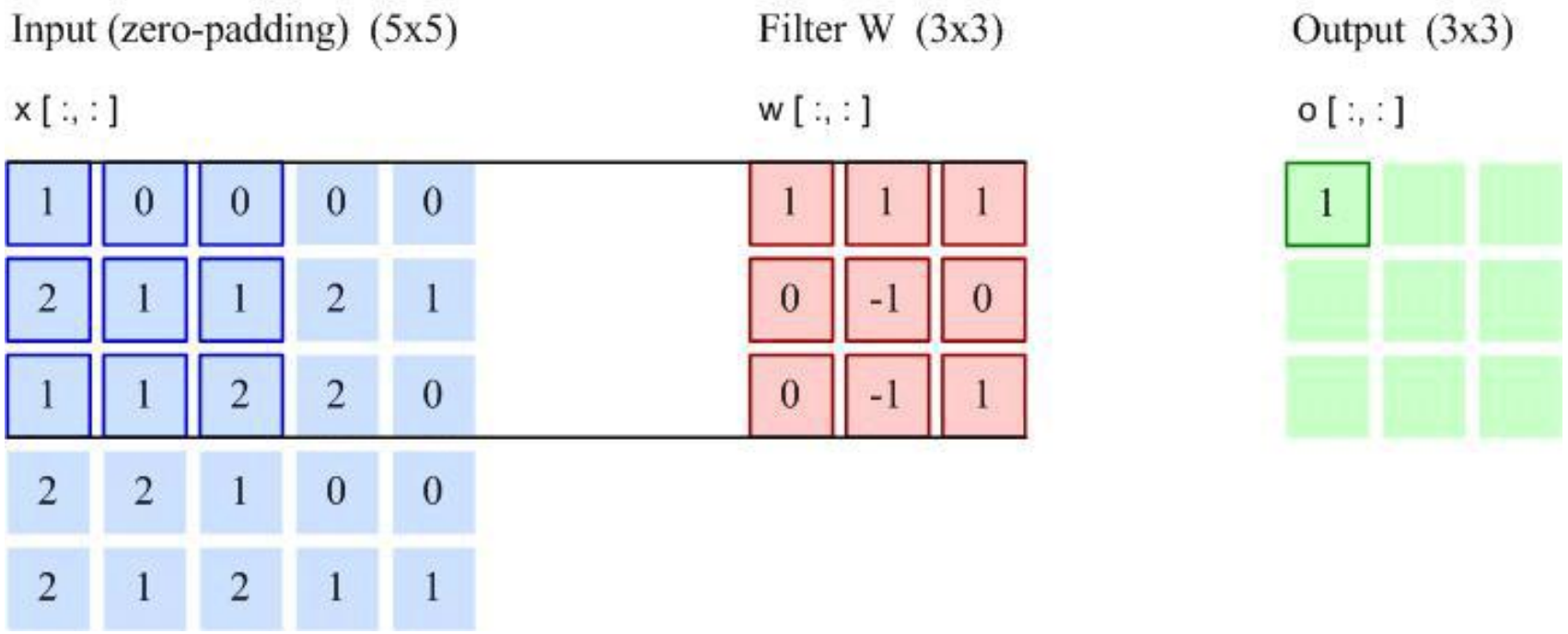


步长

- 卷积核可以移动的单位步数

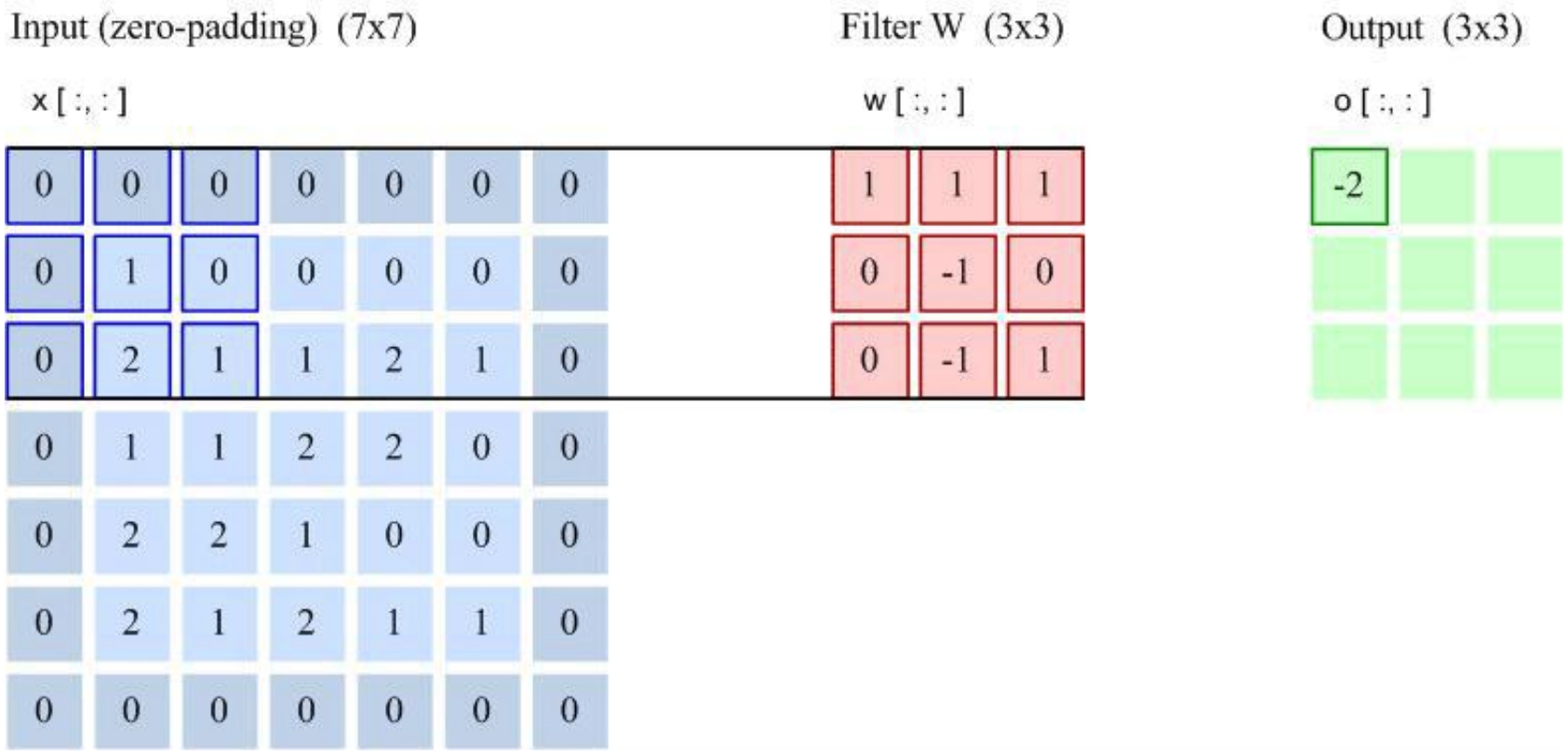
步长

- 卷积核可以移动的单位步数



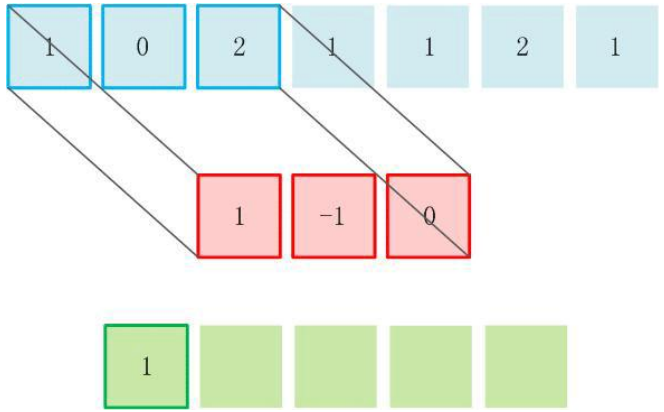
步长

- 卷积核可以移动的单位步数



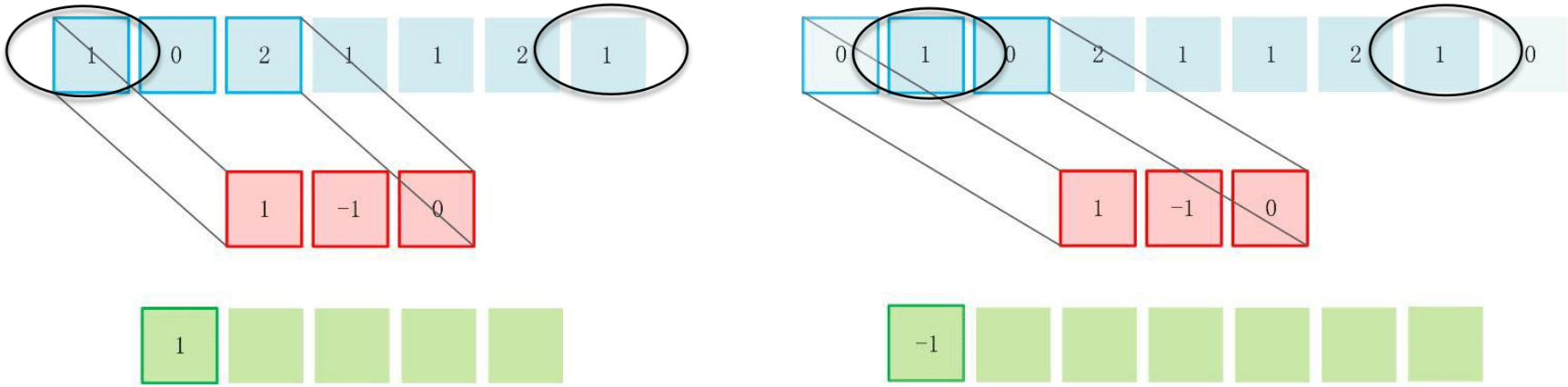
Padding

- 对输入信号边界的填充处理



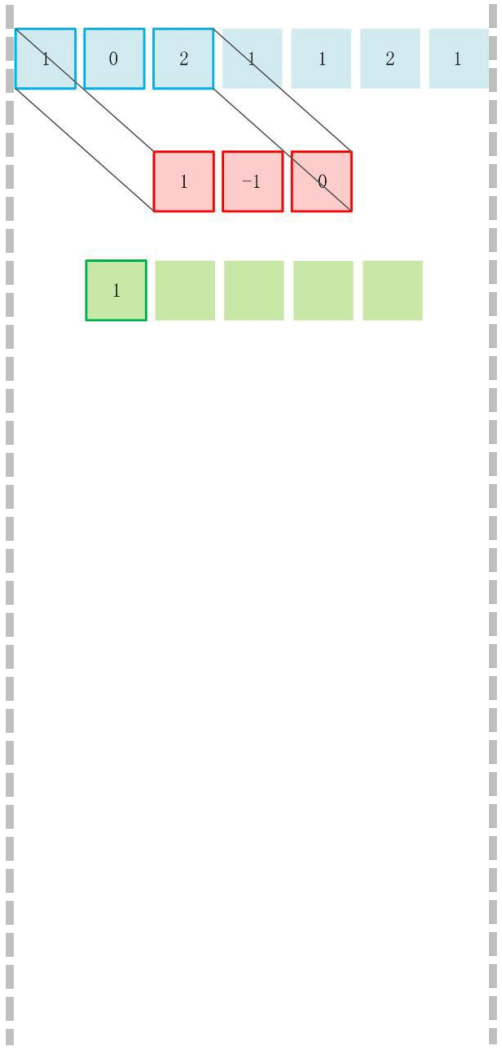
Padding

- 对输入信号边界的填充处理



三种类型卷积

Narrow



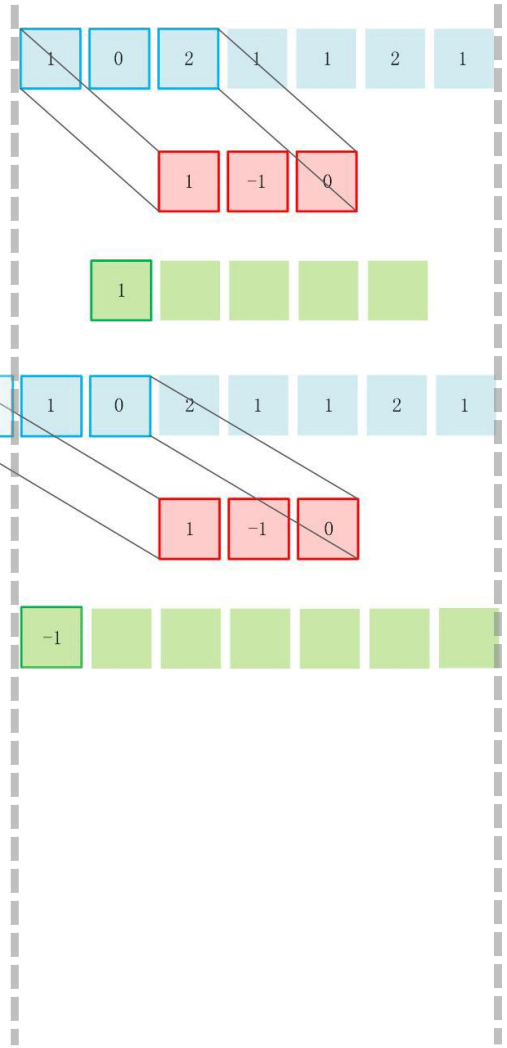
m=7

n=3

m-n+1=5

三种类型卷积

Narrow



m=7

n=3

m-n+1=5

Equal

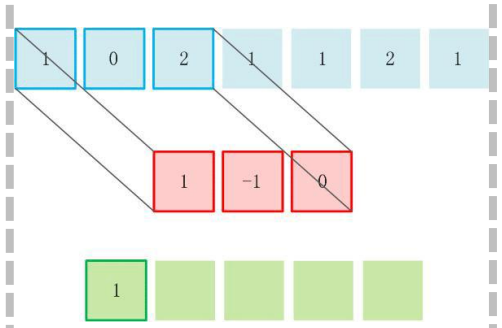
m=7

n=3

m

三种类型卷积

Narrow

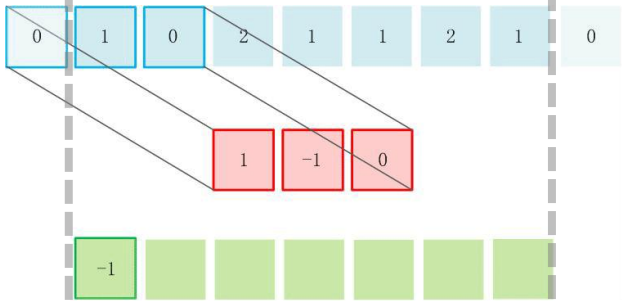


m=7

n=3

m-n+1=5

Equal

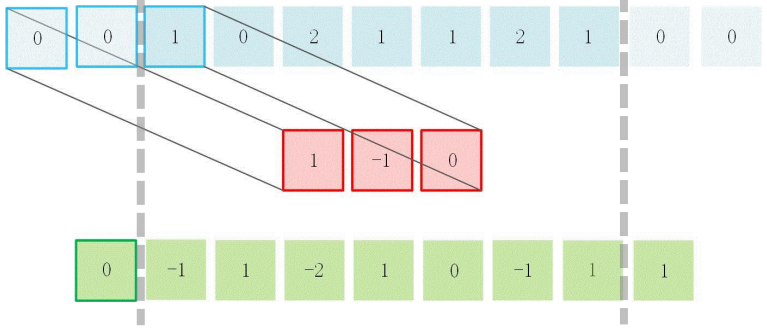


m=7

n=3

m

Wide



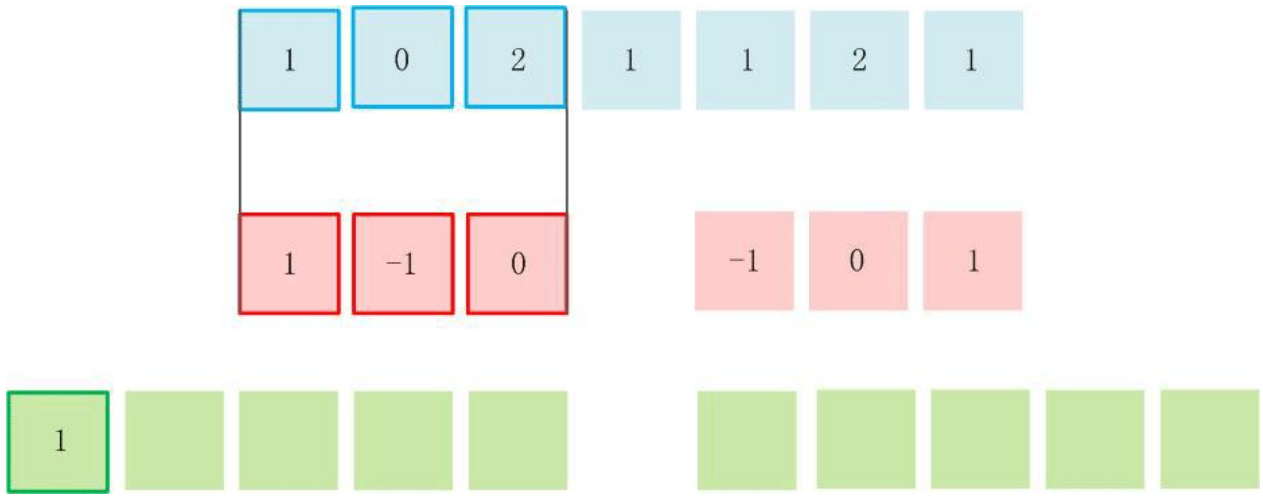
m=7

n=3

m+n-1=9

多卷积核

- 同一个输入，多个卷积进行处理



Pooling

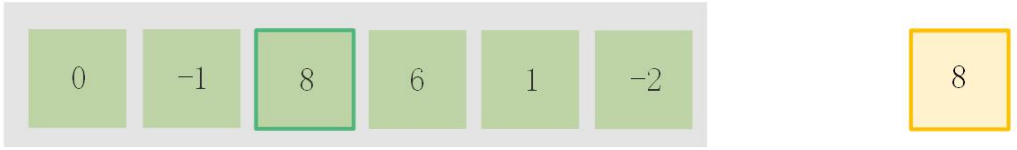
- 池化是一种聚合操作，目的是选择有信息量的特征

Pooling

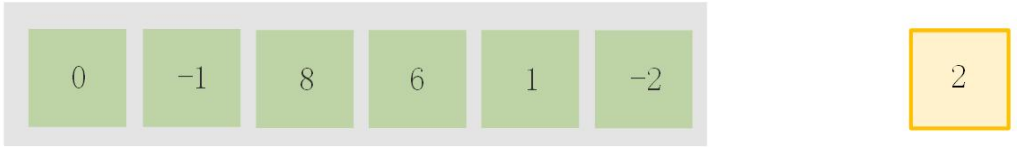
- 池化是一种聚合操作，目的是选择有信息量的特征
  - 最大池化：「你在这个范围内看到这个特征了吗？」（最常见）
  - 平均池化：「这个特征在整个范围内有多普遍？」
  - k-最大池化：「你看到这个特征多达k次了吗？」
  - 动态池化：「你在开始时看到这个特征了吗？在中间呢？在结束时呢？」

Pooling

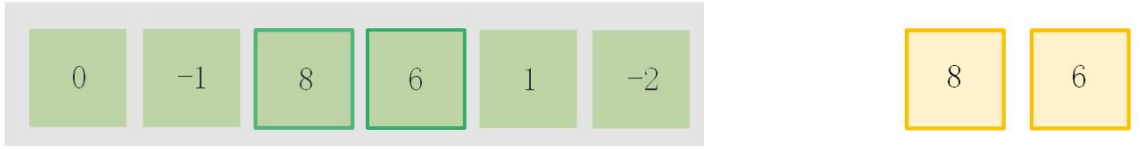
Max pooling:



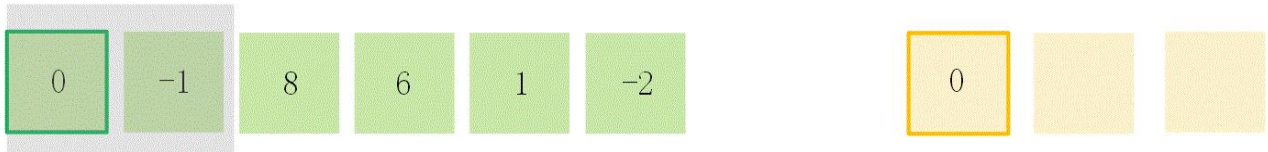
Mean pooling:



K-max pooling



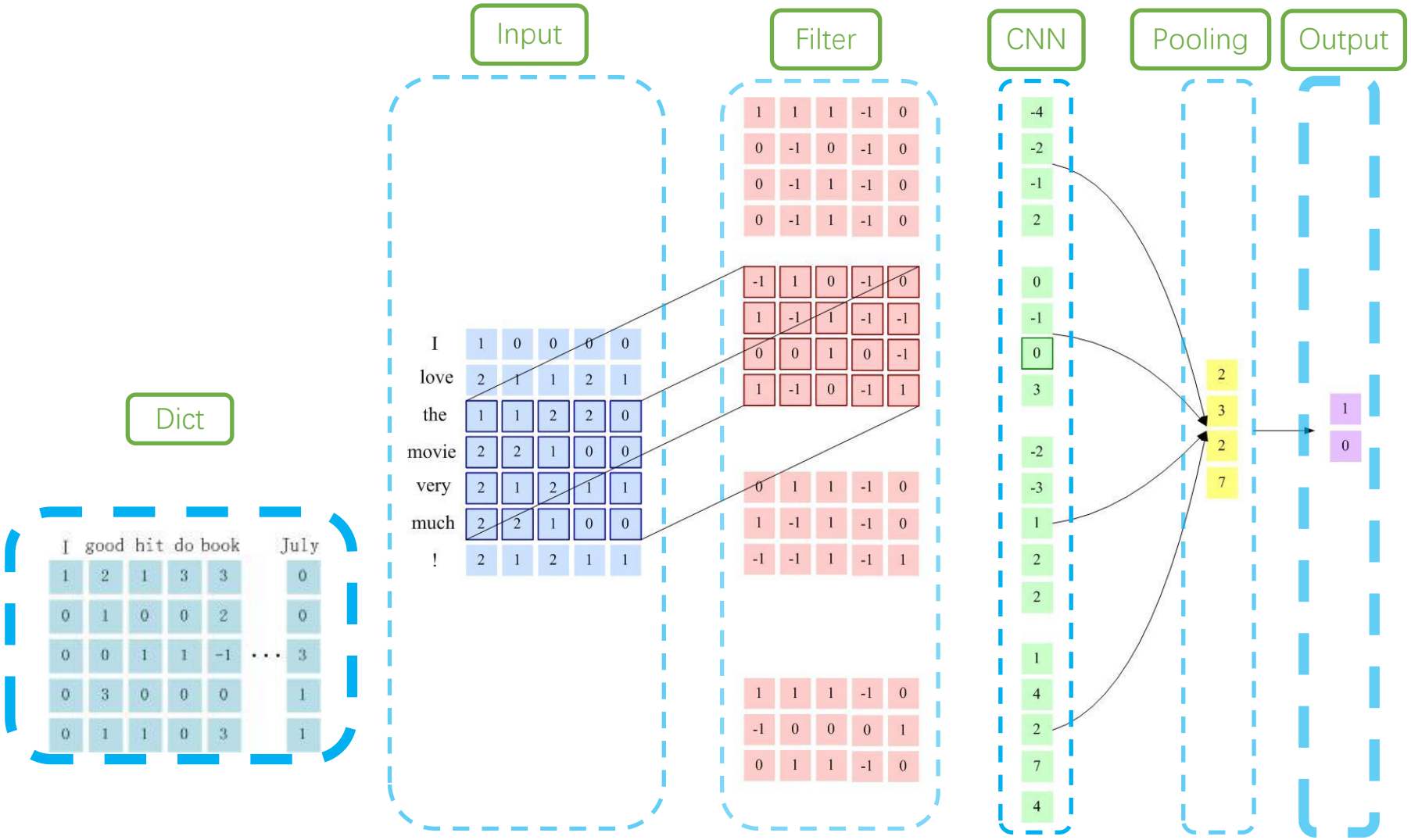
Dynamic pooling:



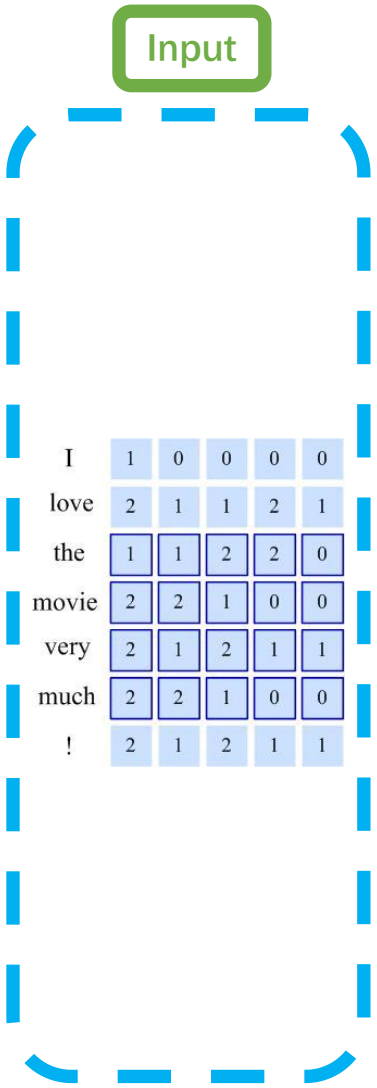
例子： CNN for Text Classification

- 任务: 情感分类
  - 输入: 句子
  - 输出: 类别 (positive/negative)
- 模型:
  - 输入层
  - 多通道卷积层
  - Pooling层、输出层

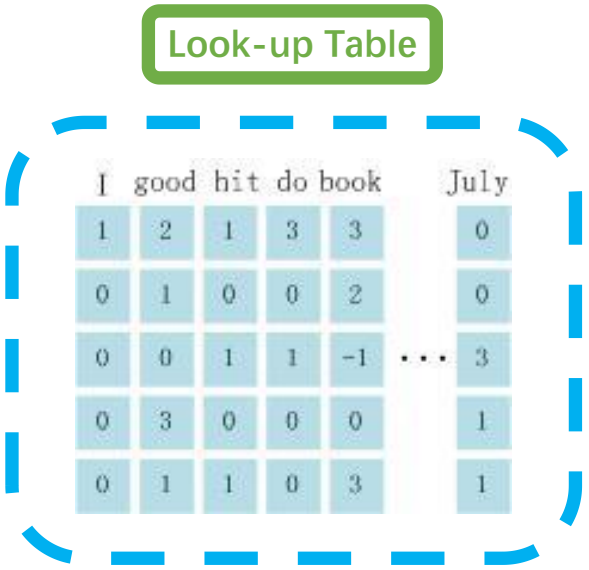
模型架构


Input
Filter
CNN
Pooling
Output
Dict
I good hit do book July
I
love
the
movie
very
much
!

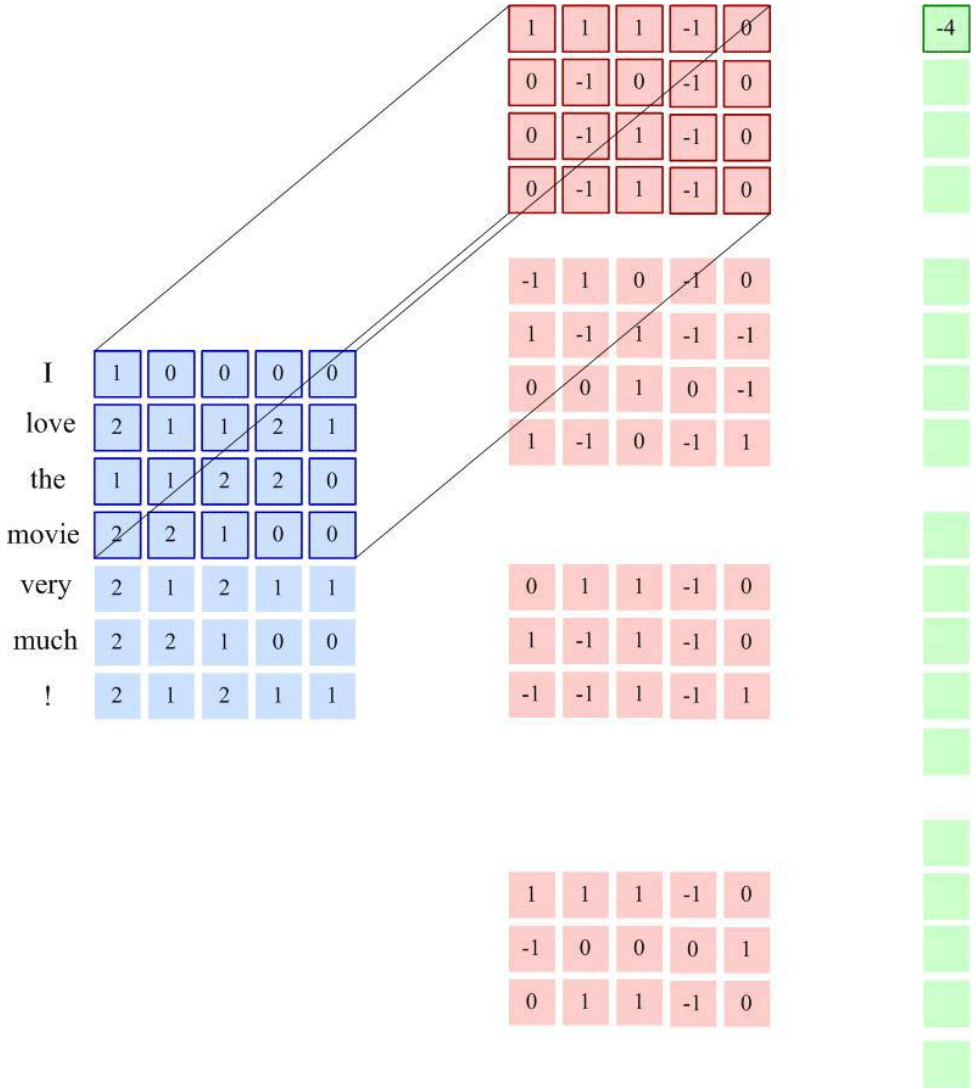
输入层



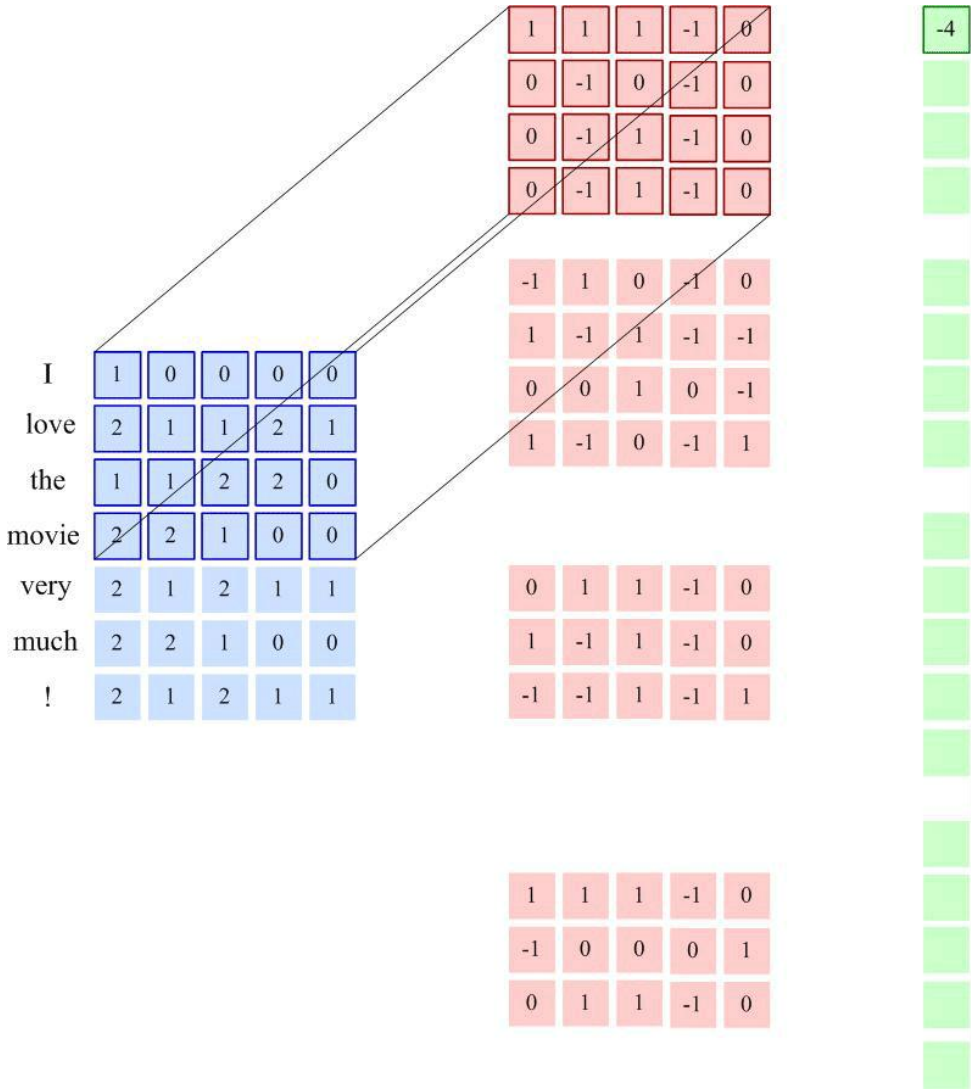
- Build a look-up table (pre-trained? Fine-tuned?)
- Discrete → distributed



卷积层

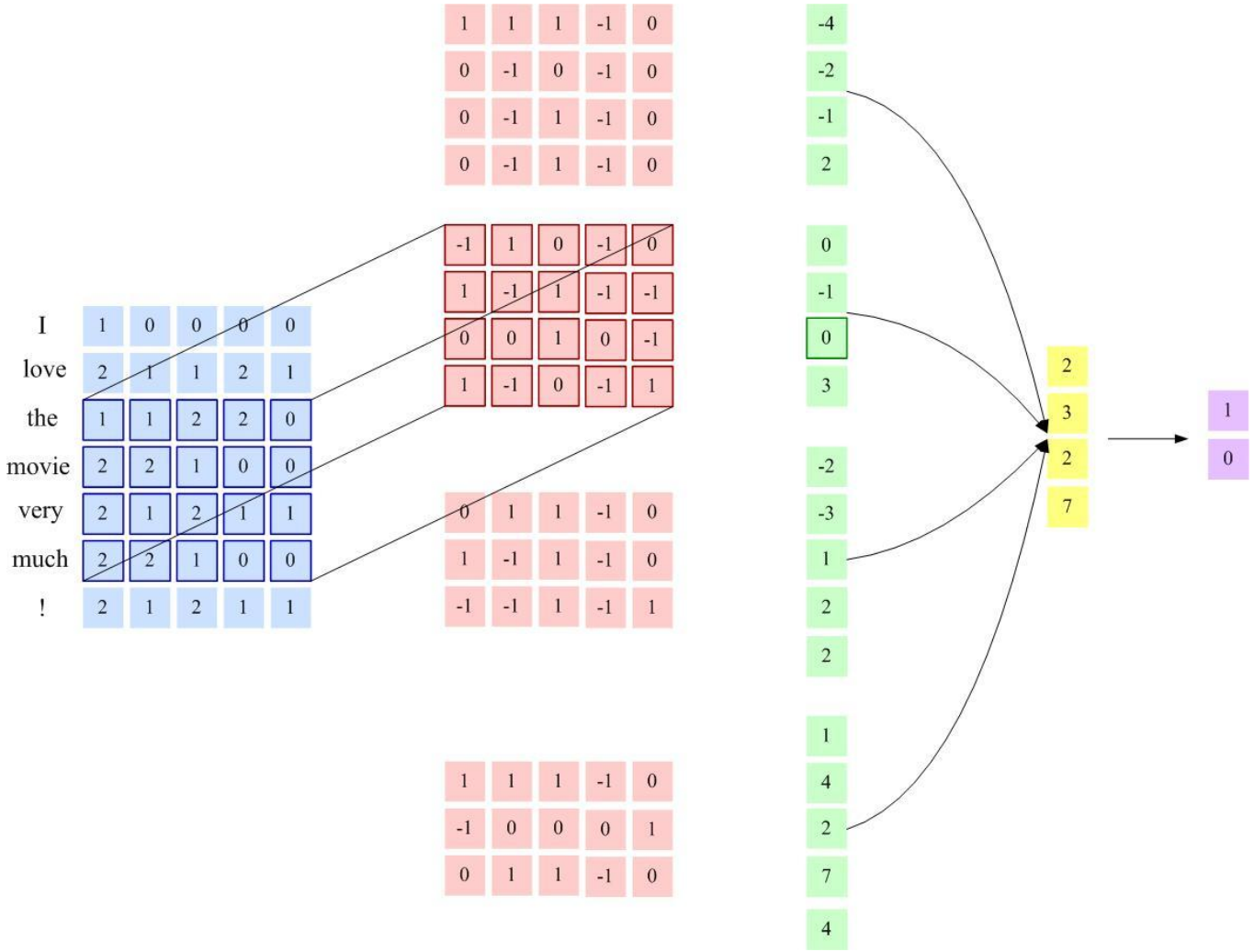

I 1 0 0 0 0
love 2 1 1 2 1
the 1 1 2 2 0
movie 2 2 1 0 0
very 2 1 2 1 1
much 2 2 1 0 0
! 2 1 2 1 1
1 1 1 -1 0
0 -1 0 -1 0
0 -1 1 -1 0
0 -1 1 -1 0
-1 1 0 -1 0
1 -1 1 -1 -1
0 0 1 0 -1
1 -1 0 -1 1
0 1 1 -1 0
1 -1 1 -1 0
-1 -1 1 -1 1
1 1 1 -1 0
-1 0 0 0 1
0 1 1 -1 0
-4

卷积层



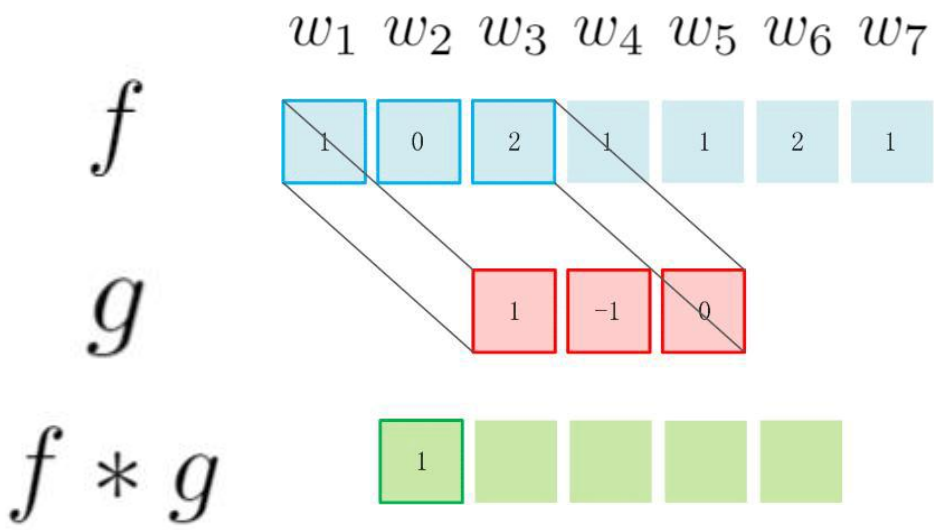
- Stride size? 1
- Wide, equal, narrow? Narrow
- How many filters? 4
- Max-pooling, Concatenate

输出层


I 1 0 0 0 0
love 2 1 1 2 1
the 1 1 2 2 0
movie 2 2 1 0 0
very 2 1 2 1 1
much 2 2 1 0 0
! 2 1 2 1 1
1 1 1 -1 0
0 -1 0 -1 0
0 -1 1 -1 0
0 -1 1 -1 0
-1 1 0 -1 0
1 -1 1 -1 -1
0 0 1 0 -1
1 -1 0 -1 1
0 1 1 -1 0
1 -1 1 -1 0
-1 -1 1 -1 1
1 1 1 -1 0
-1 0 0 0 1
0 1 1 -1 0
-4
-2
-1
2
0
-1
0
3
-2
-3
1
2
2
1
4
2
7
4
2
3
2
7
1
0

卷积神经网络的变体

CNN中的“先验”

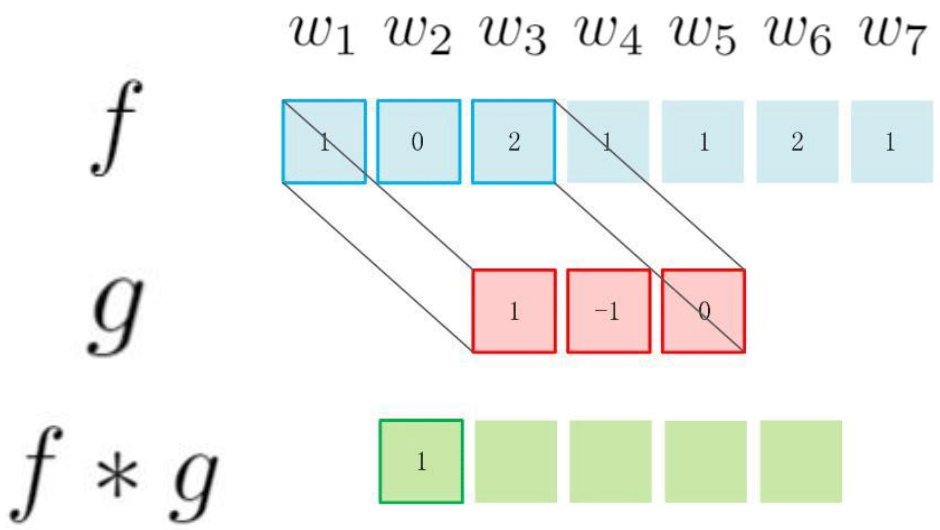


- **局部假设**
- **参数共享**

如何处理长距离依赖?

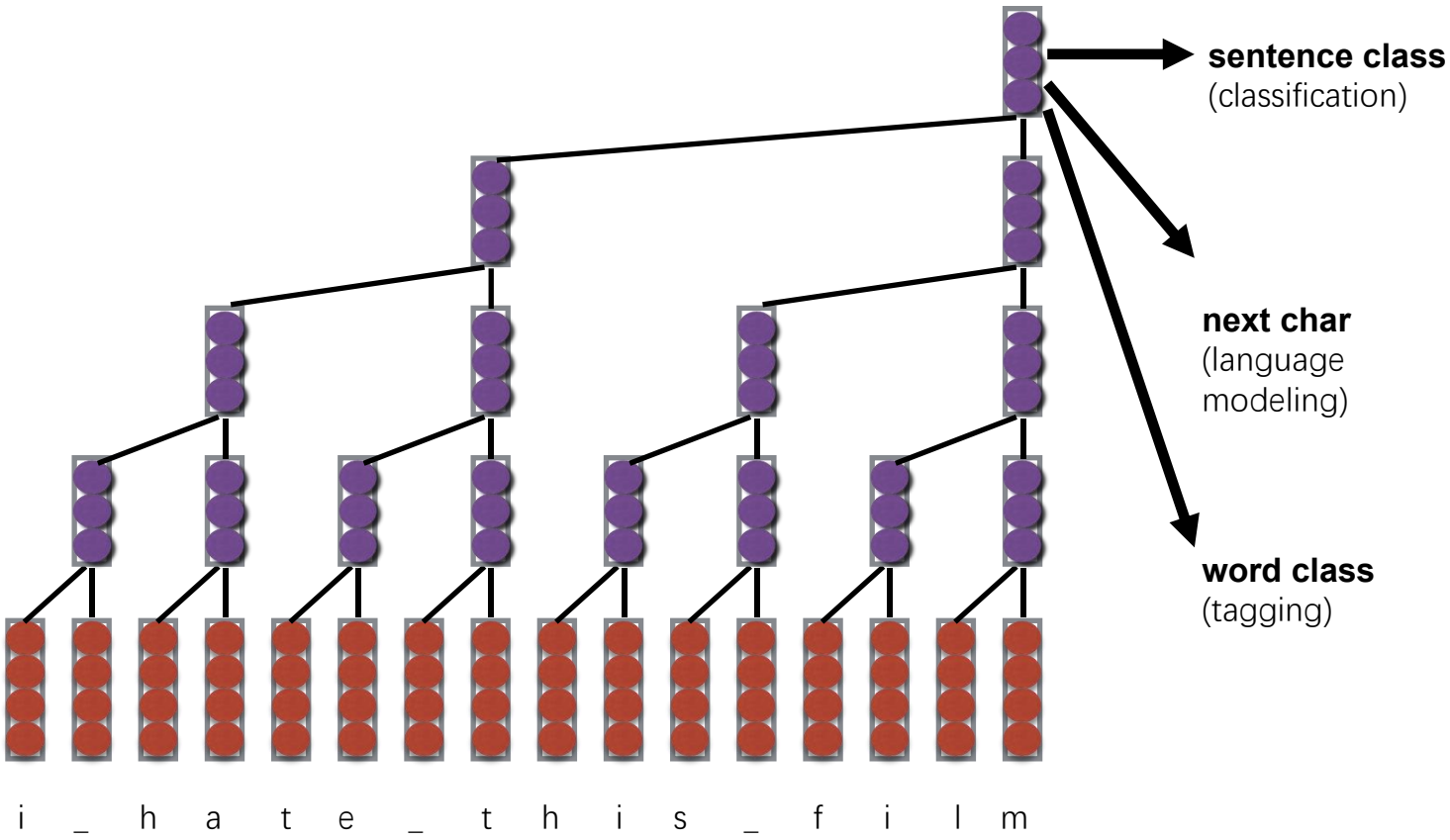
如何捕捉复杂的语义合成?

CNN中的“先验”

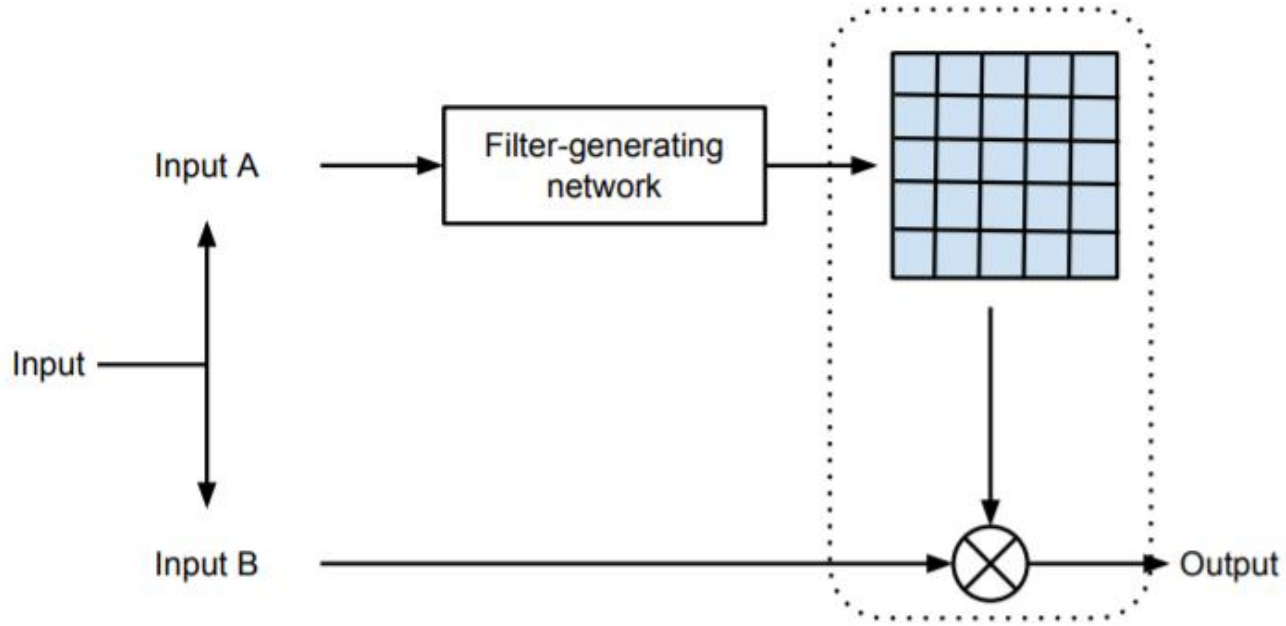


- **局部假设** 如何处理长距离依赖? **增加感受野**
- **参数共享** 如何捕捉复杂的语义合成? **动态卷积核**

Dilated Convolution (e.g. Kalchbrenner et al. 2016)

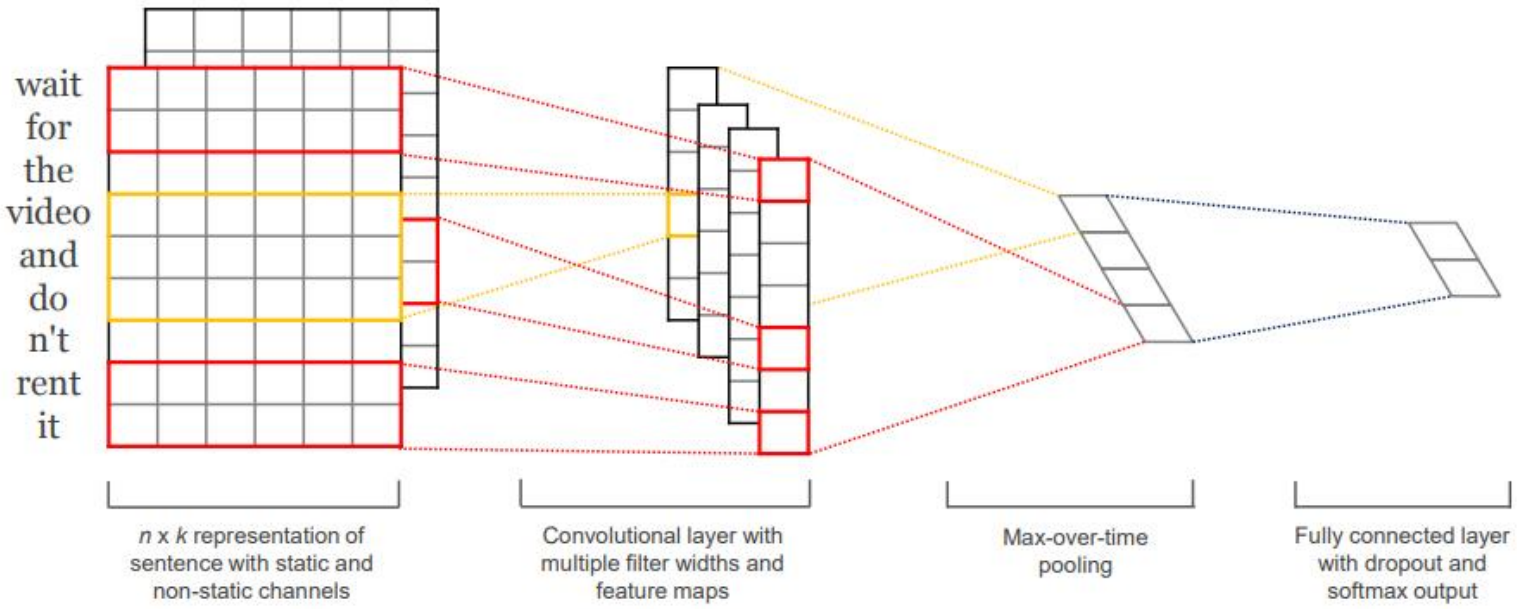


Dynamic Filter CNN (e.g. Brabandere et al. 2016)



应用

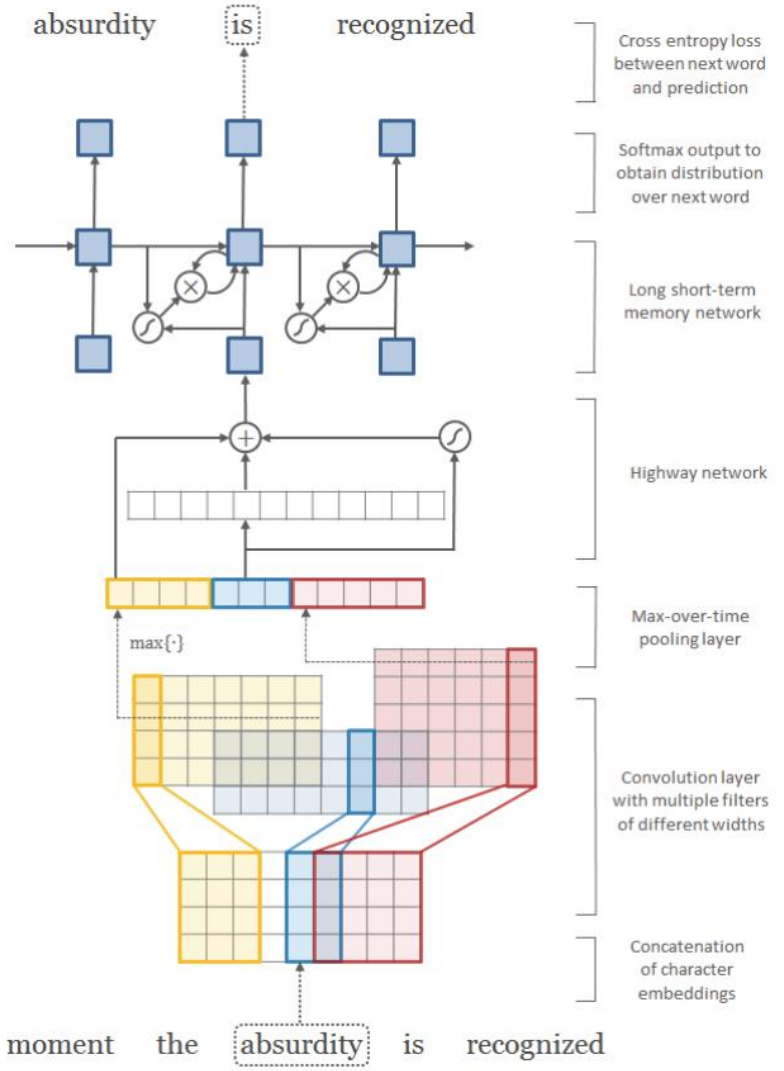
- 词级别卷积神经网络（Word-level CNNs）
  - 基本单位：词
  - 学习句子的表示
  - 短语模式



Convolutional Neural Networks for Sentence Classification, Kim et al.2014

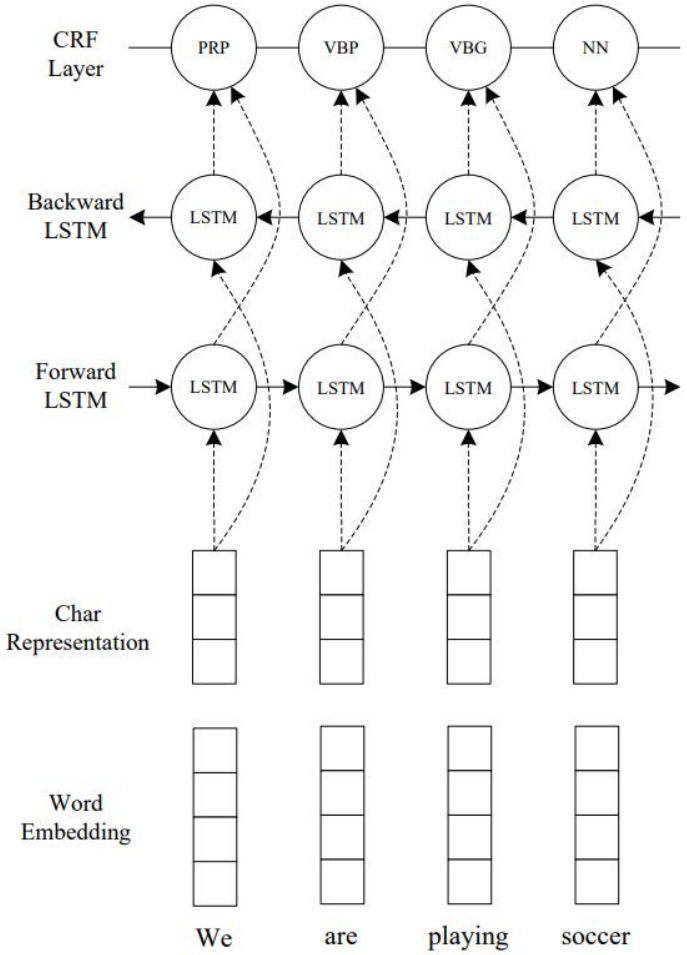
应用

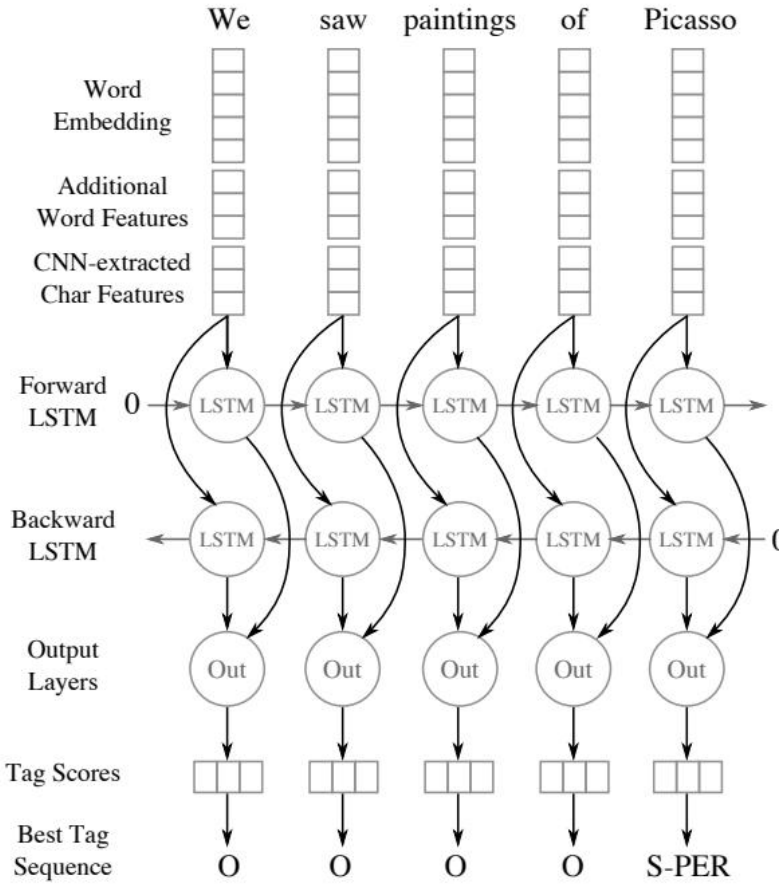
- 词级别卷积神经网络（Word-level CNNs）
- 字符级别卷积神经网络（Char-level CNNs）
  - 基本单位：字符
  - 学习单词的表示
  - 提取形态学模式



Character-Aware Neural Language Models, Kim et al 2016

多种网络混用

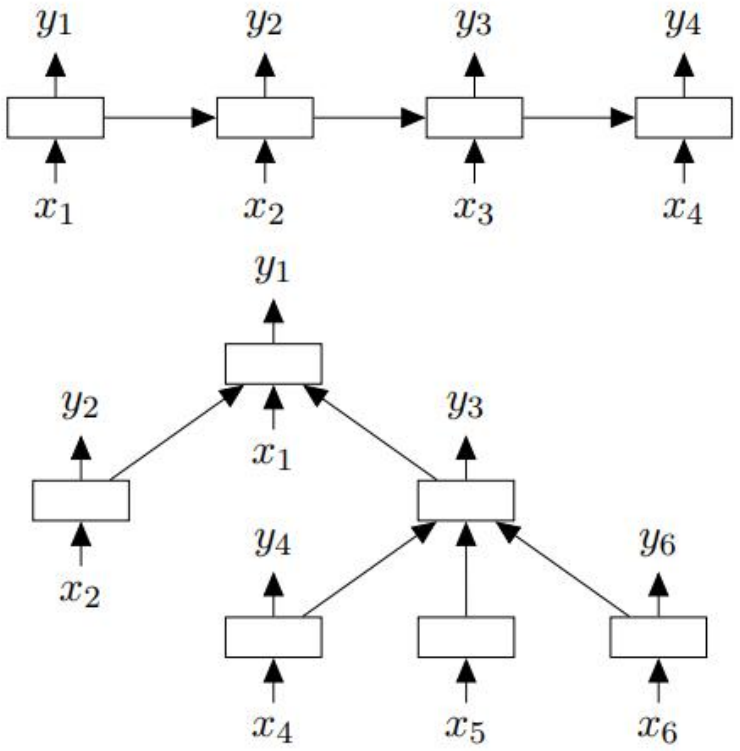




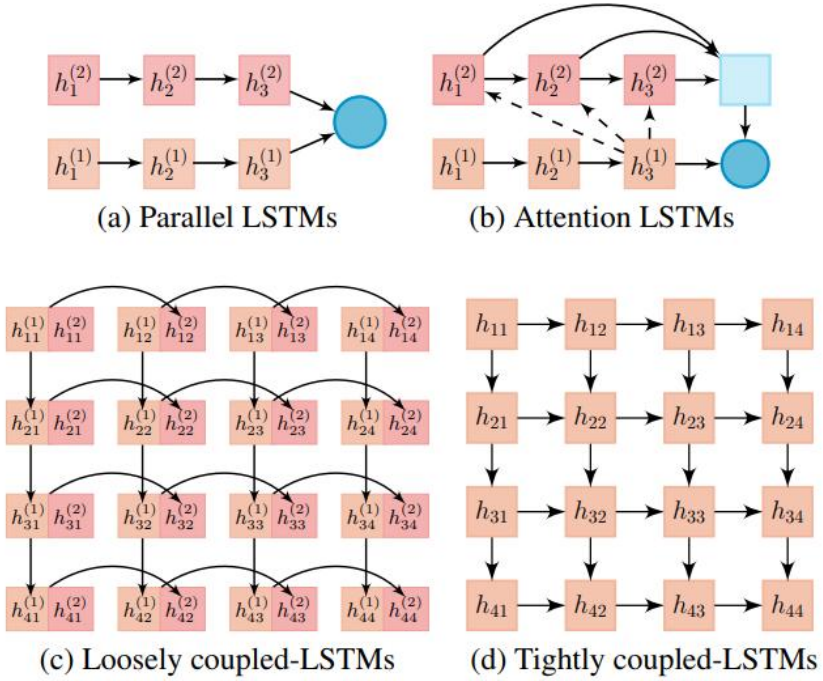
End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF Ma et al.2016

NER with Bidirectional LSTM-CNNs Chiu et al.2016

更加结构化



Tree-structure LSTM Tai et al.2015



2d LSTM Liu et al. 2016

1000+ LSTMs

小结

- 神经网络的基本概念
  - 理解基本的神经网络概念
  - 学会简单的网络梯度推导
  - 读懂代码: nn_bp.py
- 循环神经网络(RNN)
  - 理解普通RNN、LSTM的优缺点
  - 理解梯度弥散以及缓解方法
  - 了解RNN发展史和常见应用
- 卷积神经网络(CNN)
  - 掌握卷积常见概念
  - 理解CNN在NLP上的应用方法

谢谢各位!