

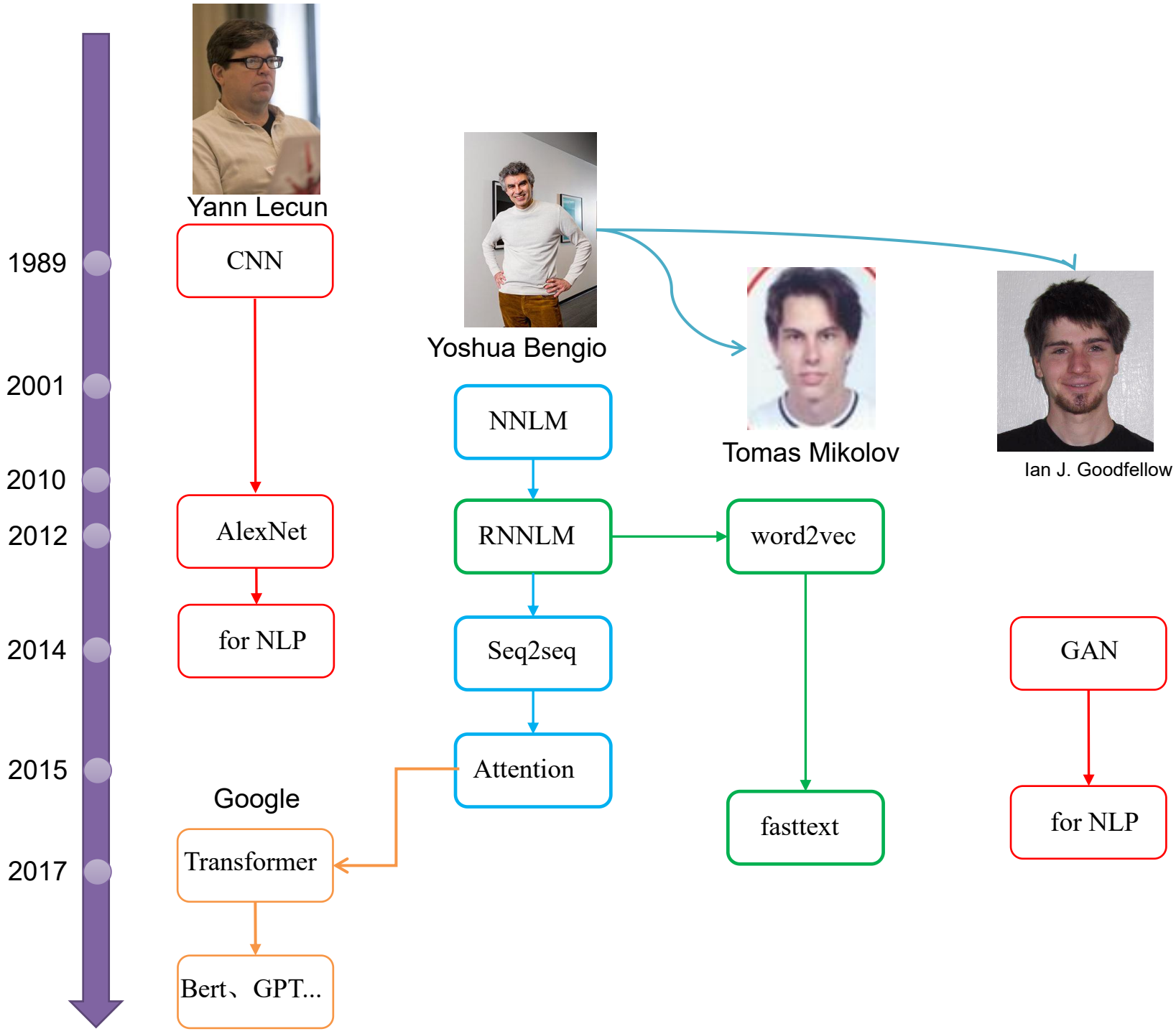
2021年度暑期强化课程

# 卷积神经网络CNN

**授课人：曹亚男**



中国科学院 信息工程研究所  
INSTITUTE OF INFORMATION ENGINEERING, CAS



# 7. 卷积神经网络CNN

---

7.1

CNN模型简介

7.2

CNN的超参数

7.3

CNN for NLP

# 7. 卷积神经网络CNN

---

7.1

CNN模型简介

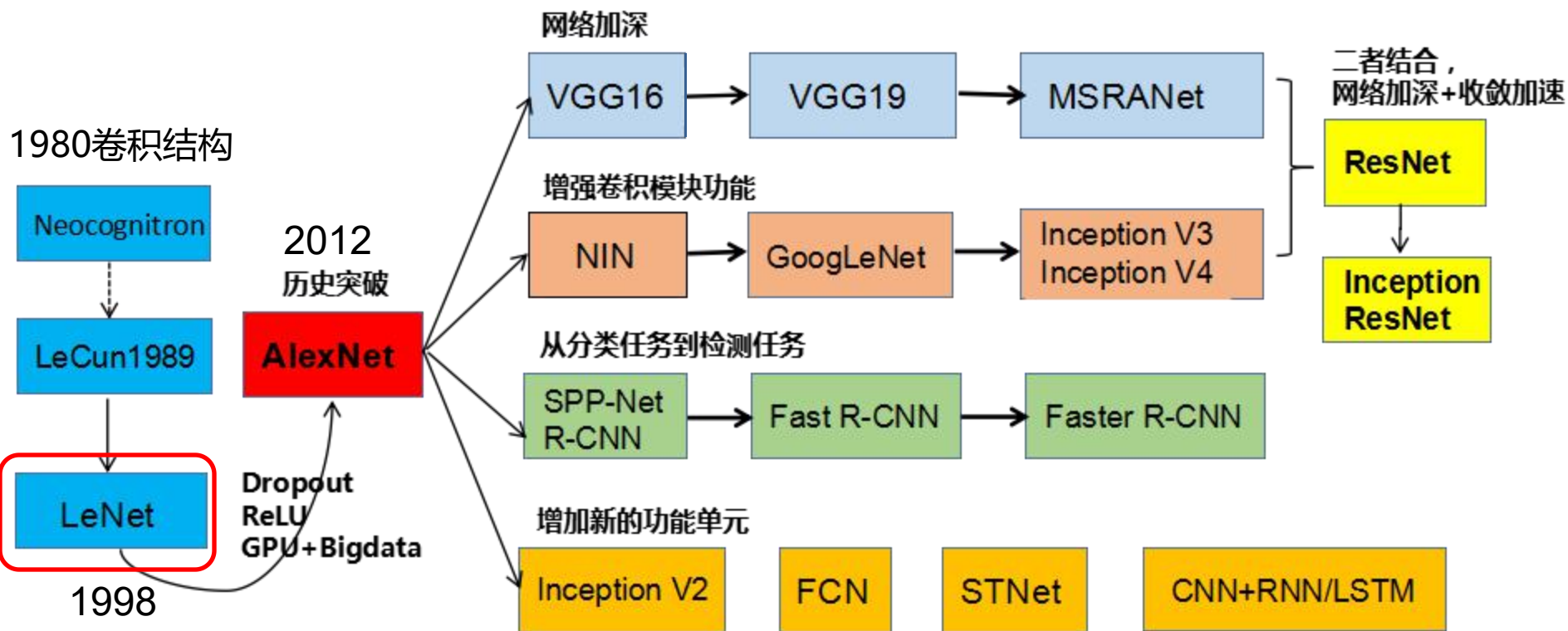
7.2

CNN的超参数

7.3

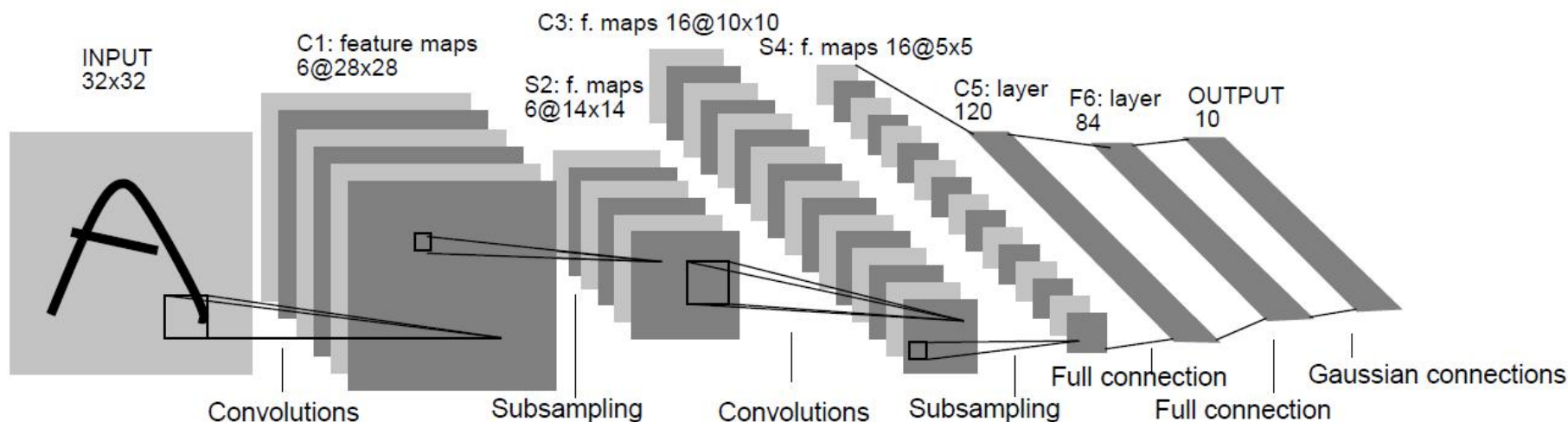
CNN for NLP

# CNN模型的演化历程



# 基础模型结构

## 最早的CNN: LeNet-5模型结构



- Input: 二维数据，典型的数据类型是图片，包含 $n*n$ 个像素点
- Convolution: 基于 $m*m$ 的滤波器，对输入数据进行卷积操作，输出特征图
- Subsampling: 对特征图进行降采样（池化），生成更小的特征图
- Gaussian connection: 输出层使用高斯函数，目前常用的是softmax函数

# 卷积 (Convolution)

- 卷积操作

- 二维卷积  $H(x, y) = I * K(x, y) = \sum_m \sum_n I(m, n)K(x - m, y - n)$

1	0	1
0	1	0
1	0	1

filter

1 <sub>x1</sub>	1 <sub>x0</sub>	1 <sub>x1</sub>	0	0
0 <sub>x0</sub>	1 <sub>x1</sub>	1 <sub>x0</sub>	1	0
0 <sub>x1</sub>	0 <sub>x0</sub>	1 <sub>x1</sub>	1	1
0	0	1	1	0
0	1	1	0	0

Image

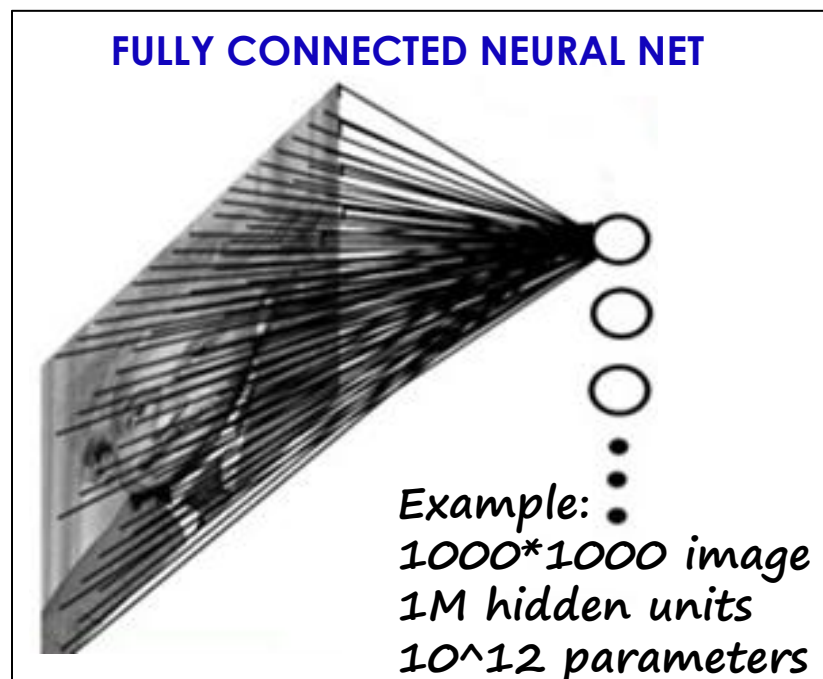
4		

Convolved  
feature

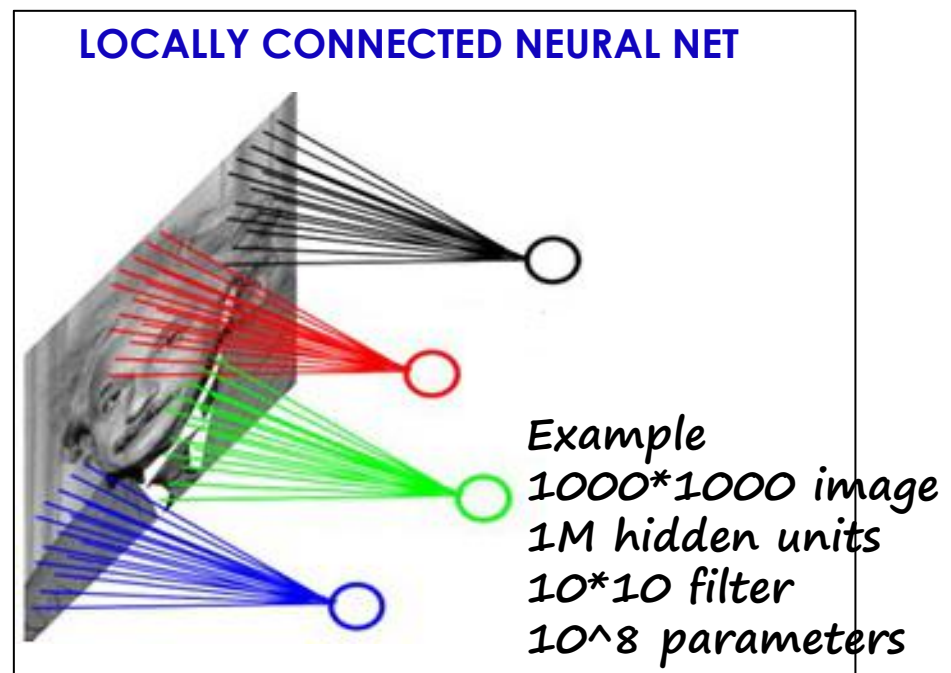
- 对于某个时刻的窗口，通过神经网络的非线性变换，将这个窗口内的输入值转换为某个特征值，随着窗口的移动，形成这个滤波器的特征向量

# 卷积层的特点 (1)

- 稀疏连接（局部连接）



传统神经网络层：全连接

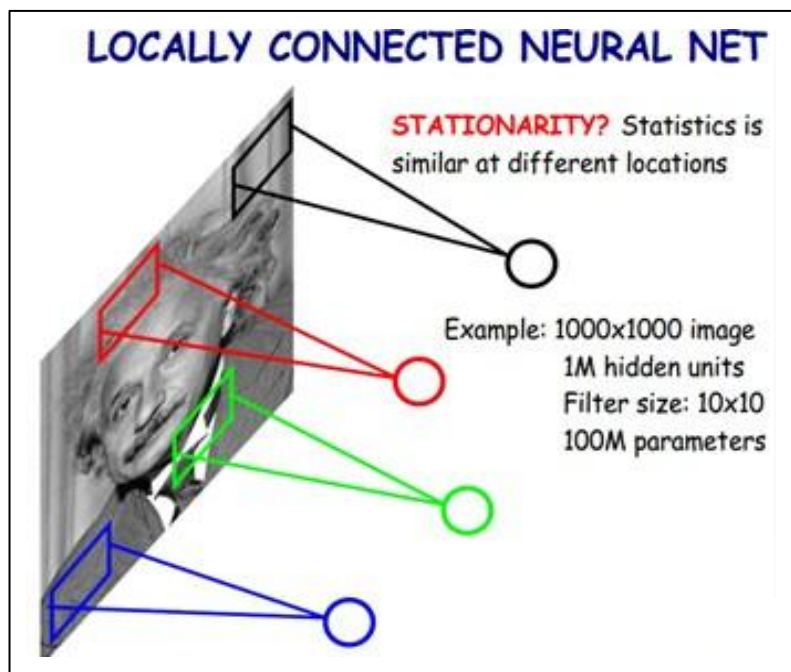


卷积层：稀疏连接

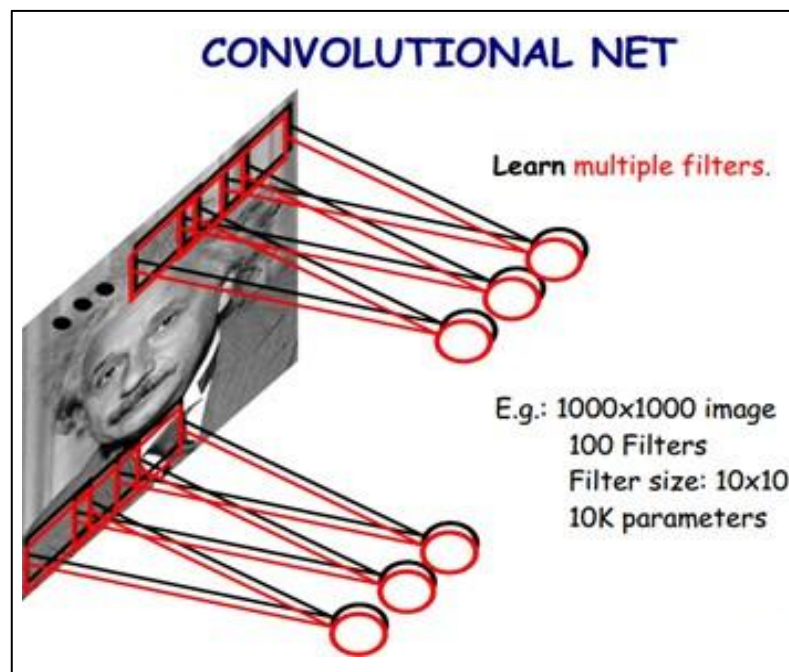


# 卷积层的特点 (2)

- 权值共享



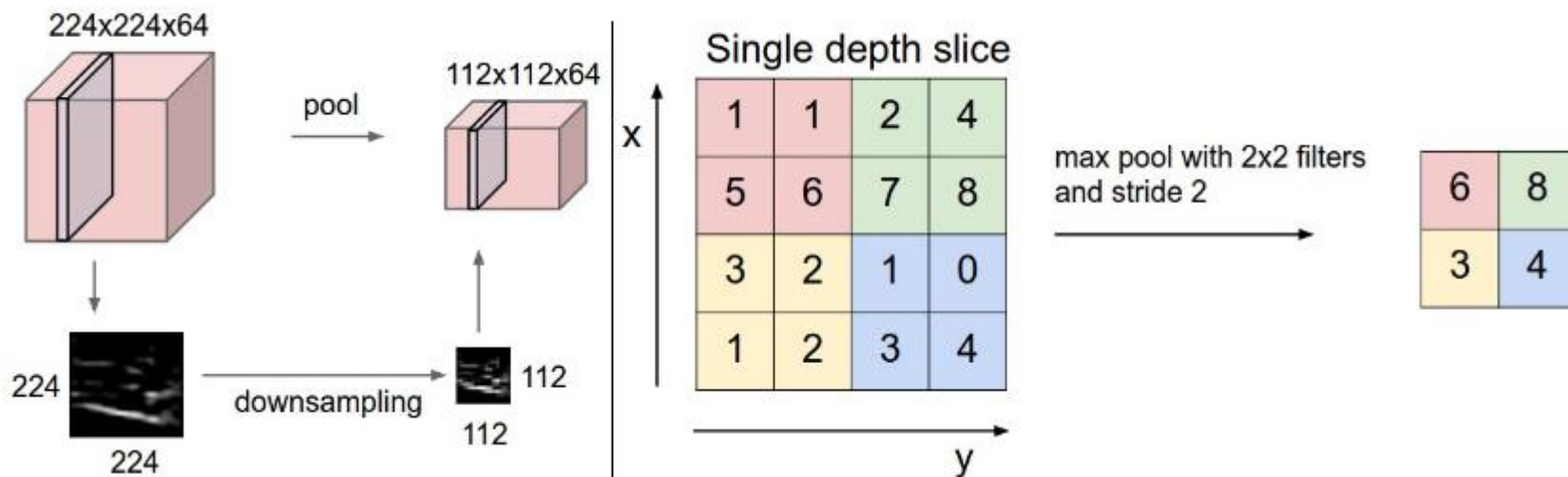
稀疏连接：不共享权值



稀疏连接：共享权值

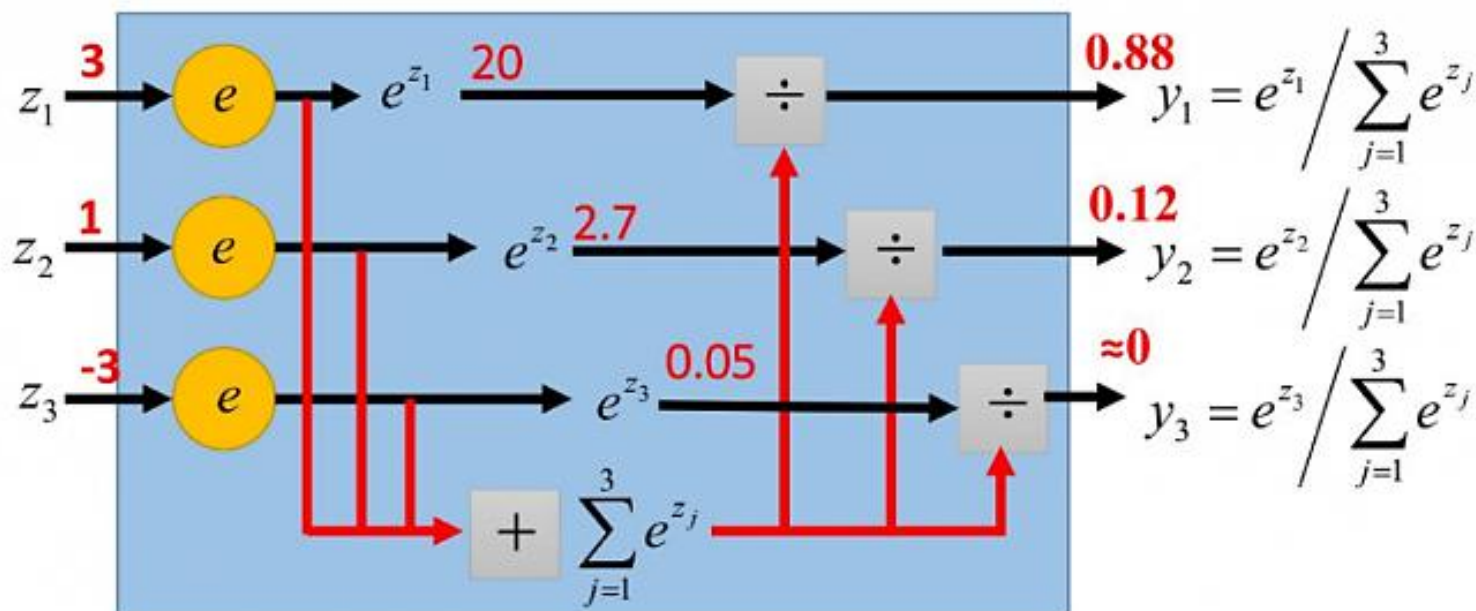
# 池化(Pooling)

- 池化：一种降采样操作，用于降低特征维度并保留有效信息
  - 减少模型参数，避免过拟合，提高训练速度
  - 保证特征的位置、旋转、伸缩不变性（CV）
  - 将变长的输入转换成固定长度（NLP）

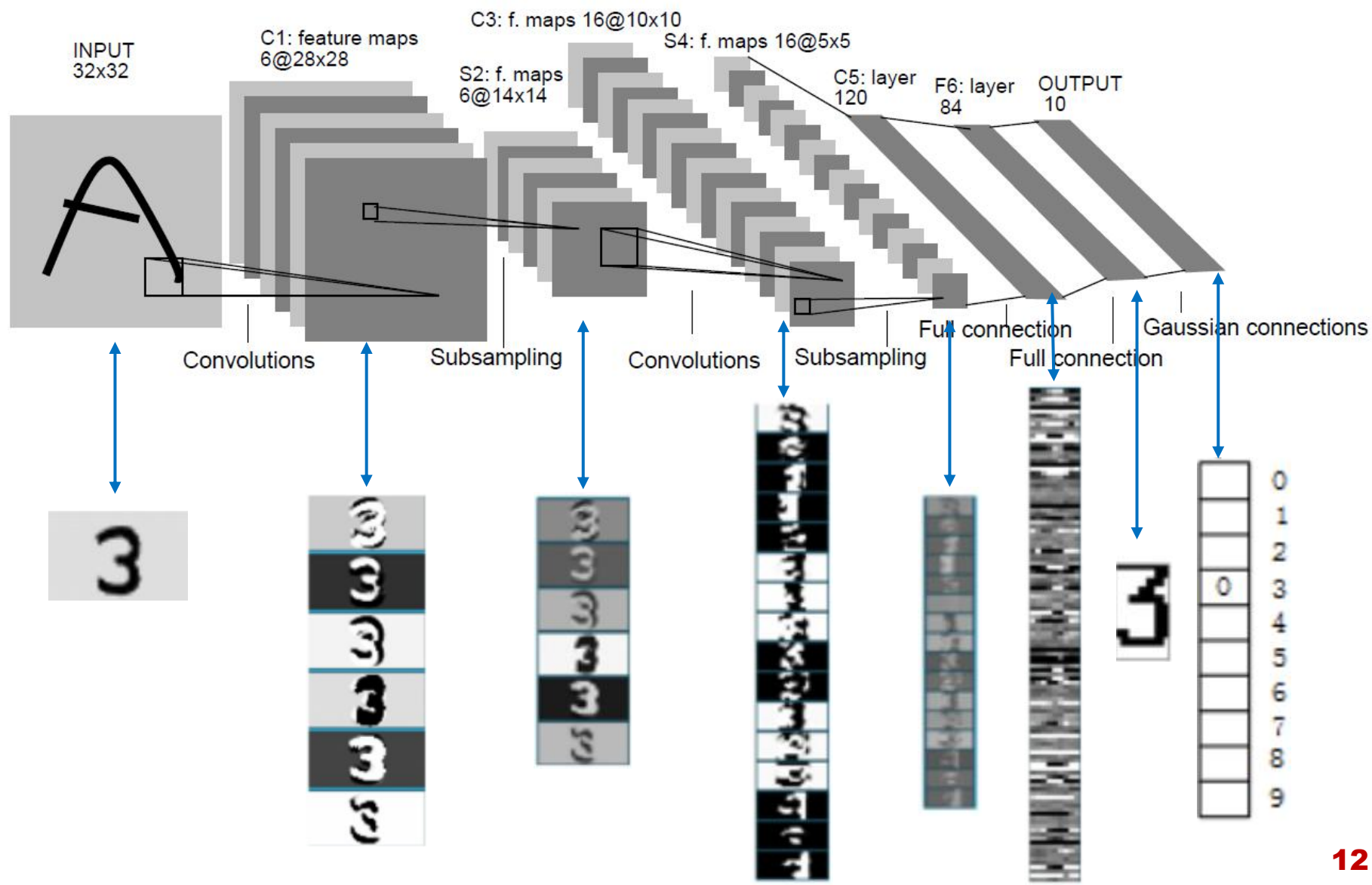


# Softmax函数

- 函数形式  $y_i = e^{z_i} / \sum_{j=1} e^{z_j}$
- 因  $0 < y_i < 1$ ，且  $\sum_i y_i = 1$ ，softmax 可以看作输出概率；  
应用在分类问题中，选择概率最大的节点，作为预测结果



# LeNet-5识别数字过程



# 7. 卷积神经网络CNN

---

7.1

CNN模型简介

7.2

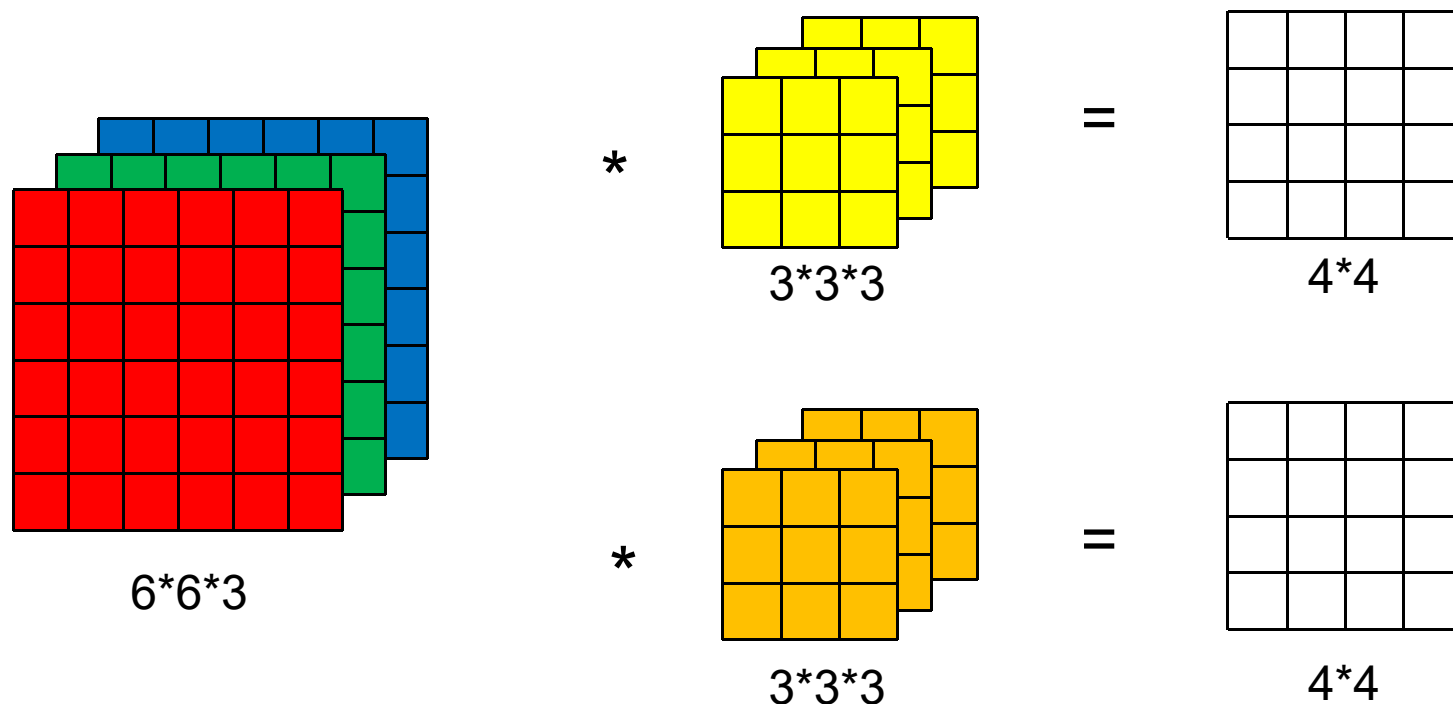
CNN的超参数

7.3

CNN for NLP

# 通道

- 输入数据通道：取决于输入数据的类型，如RGB图片的通道数是3
- 卷积操作输出的通道：取决于卷积核的数量，下图输出通道数是2



# 宽卷积 VS. 窄卷积

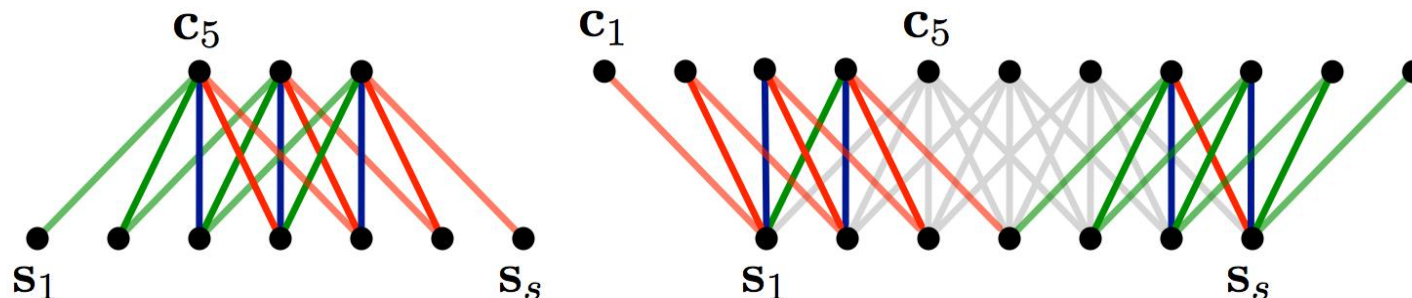
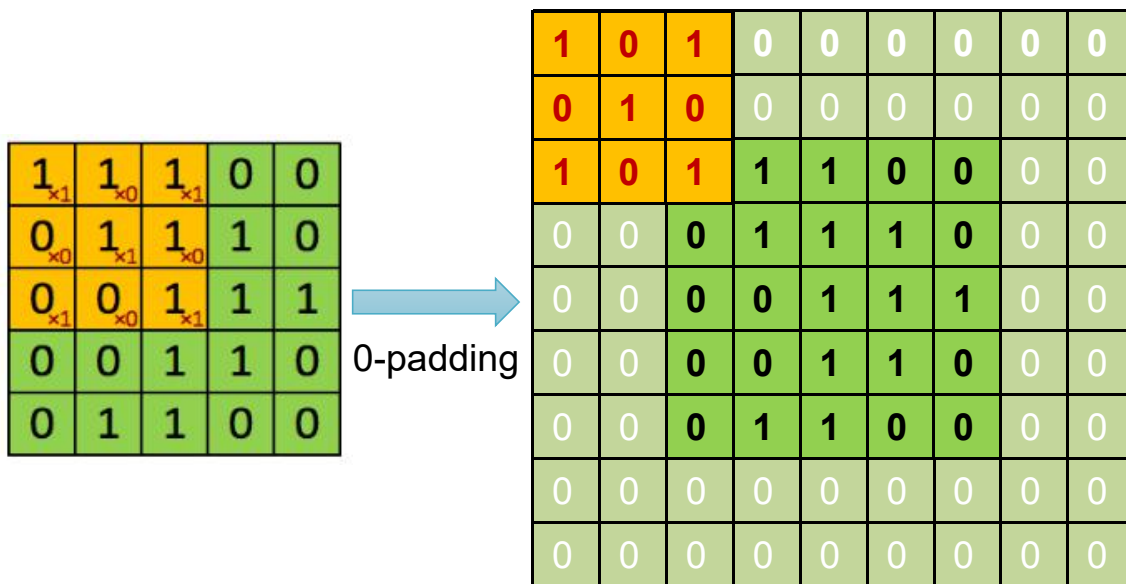


Figure. Narrow vs. Wide Convolution. Filter size 5, input size 7



窄卷积

宽卷积

Input matrix size =  $n$

Padding size =  $p$

Filter size =  $f$

Output matrix size = ?

How to make it equal to  $n$ ?

# 步长 (Stride Size)

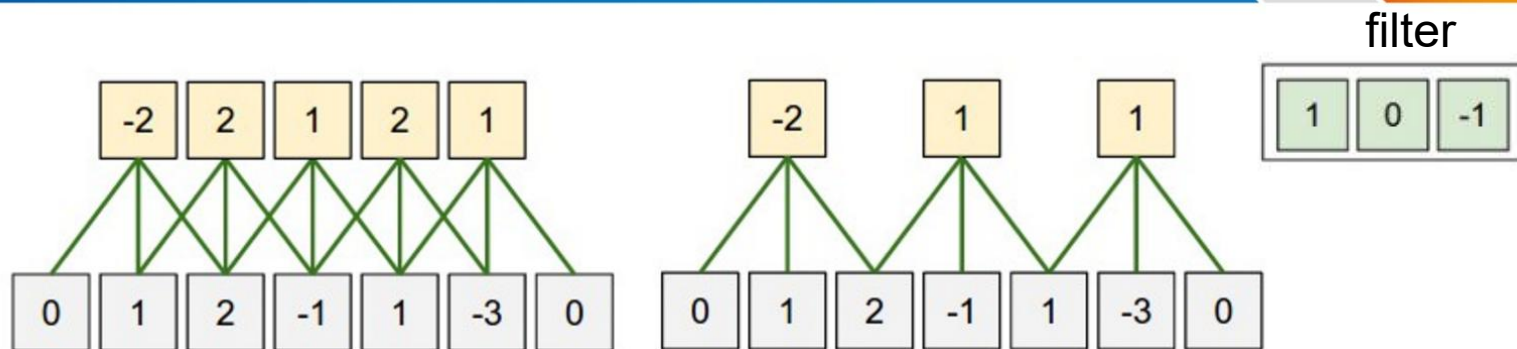


Figure. Stride size 1 vs. Stride size 2. Filter size 3, input size 7

1	0	1	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0
1	0	1	1	1	0	0	0	0
0	0	0	1	1	1	0	0	0
0	0	0	0	1	1	1	0	0
0	0	0	0	1	1	0	0	0
0	0	0	1	1	0	0	0	0
0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0

宽卷积

Input matrix size =  $n$

Padding size =  $p$

Filter size =  $f$

Stride size =  $s$

Output matrix size = ?

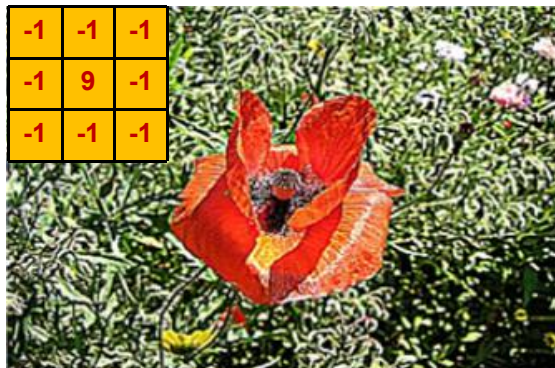
Constraints on strides?



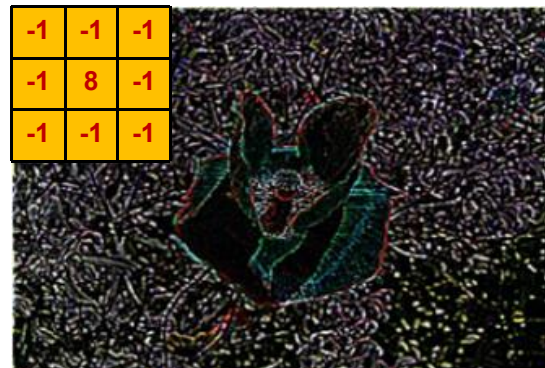
# 卷积核的数量和尺寸



原图



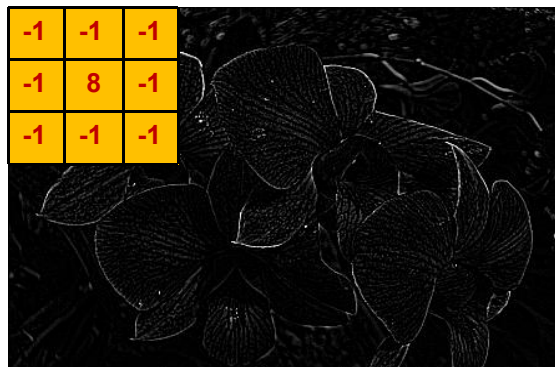
锐化 (3\*3)



边缘检测 (3\*3)



原图

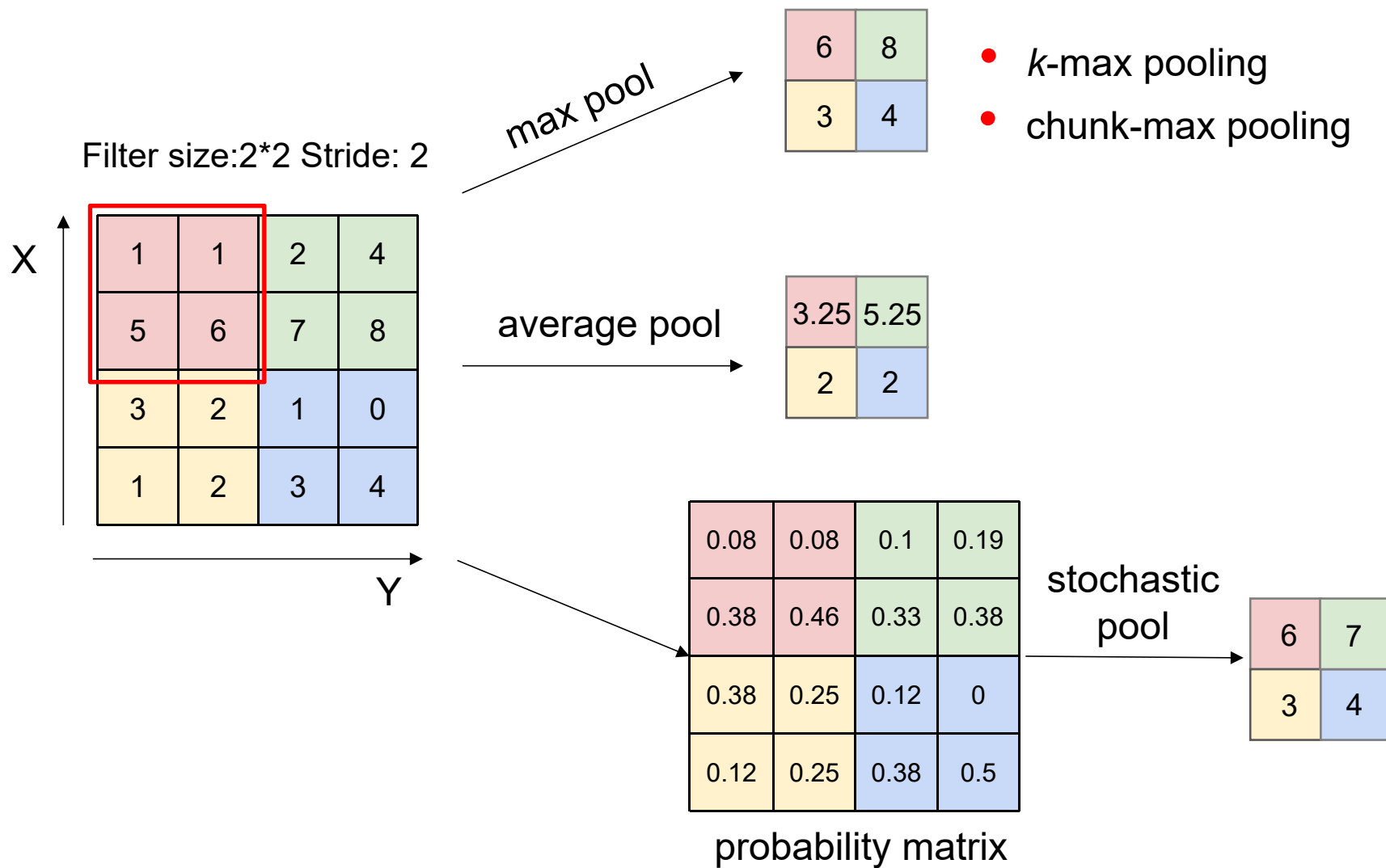


边缘检测 (3\*3)

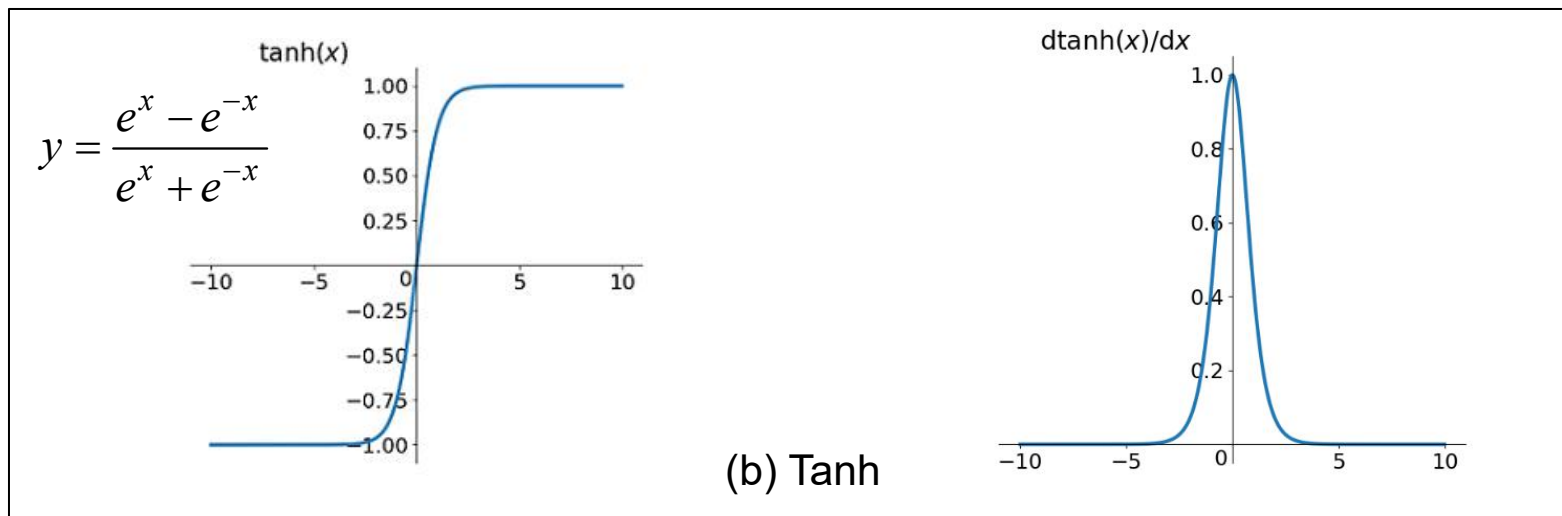
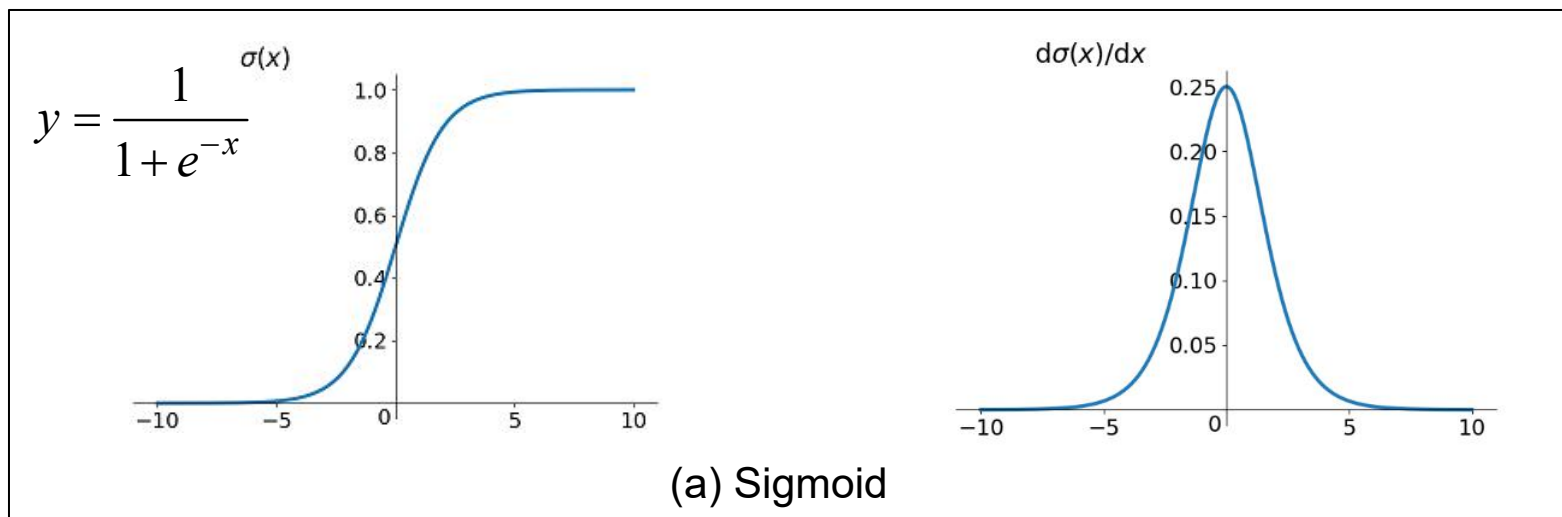


边缘检测 (5\*5)

# 各类池化



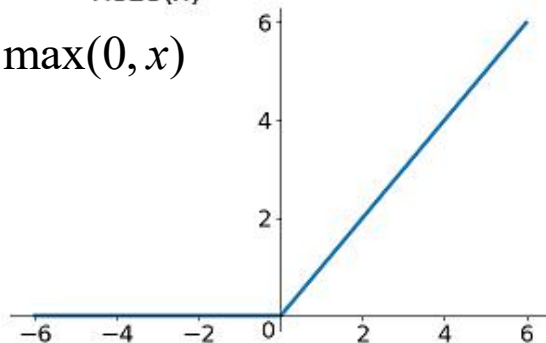
# 激活函数



# 激活函数

ReLU(x)

$$y = \max(0, x)$$

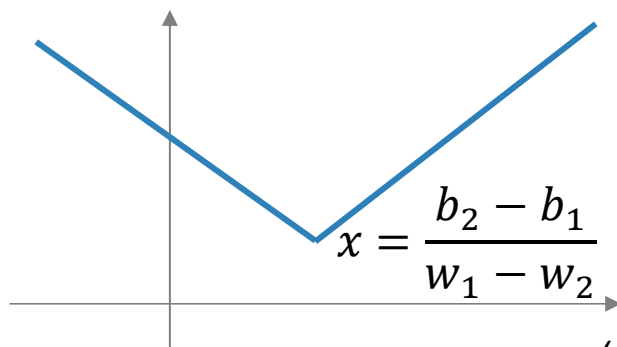


$d\text{ReLU}(x)/dx$

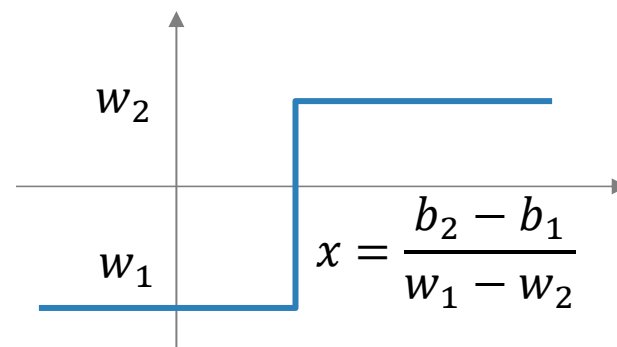


(c) ReLU

$$y = \max(w_1x + b_1, w_2x + b_2)$$



$d\text{maxout}(x) / dx$



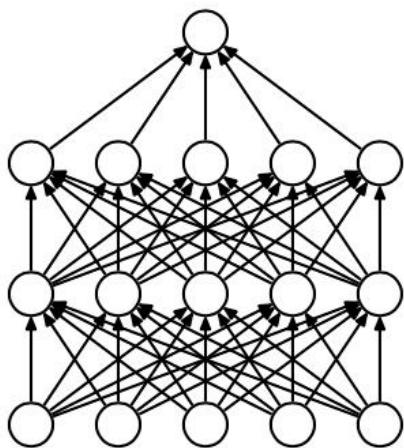
(d) Maxout

# 正则: Dropout

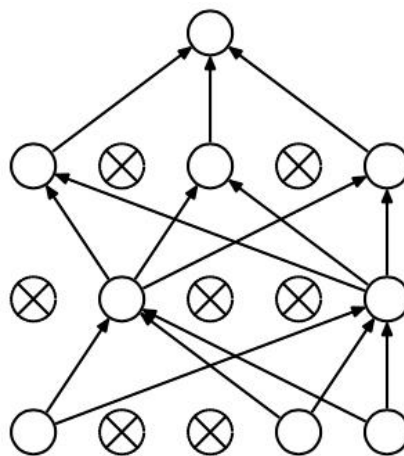
- 随机正则化策略，提高模型泛化能力，避免过拟合
  - Dropout引入Bernoulli随机数 $u$ ， $p$ 代表dropout ratio，表示在训练的前向传导中，让某个神经元激活值为0的概率

$$y_{train} = \begin{cases} \frac{x}{1-p} & \text{if } u > p \\ 0 & \text{otherwise} \end{cases} \quad \text{where, } u \sim U(0,1)$$

$$E(y_{train}) = p \cdot 0 + (1-p) \frac{E(x)}{1-p} = E(x)$$



(a) Standard NN



(b) After applying dropout

# 如何设置超参数?

---

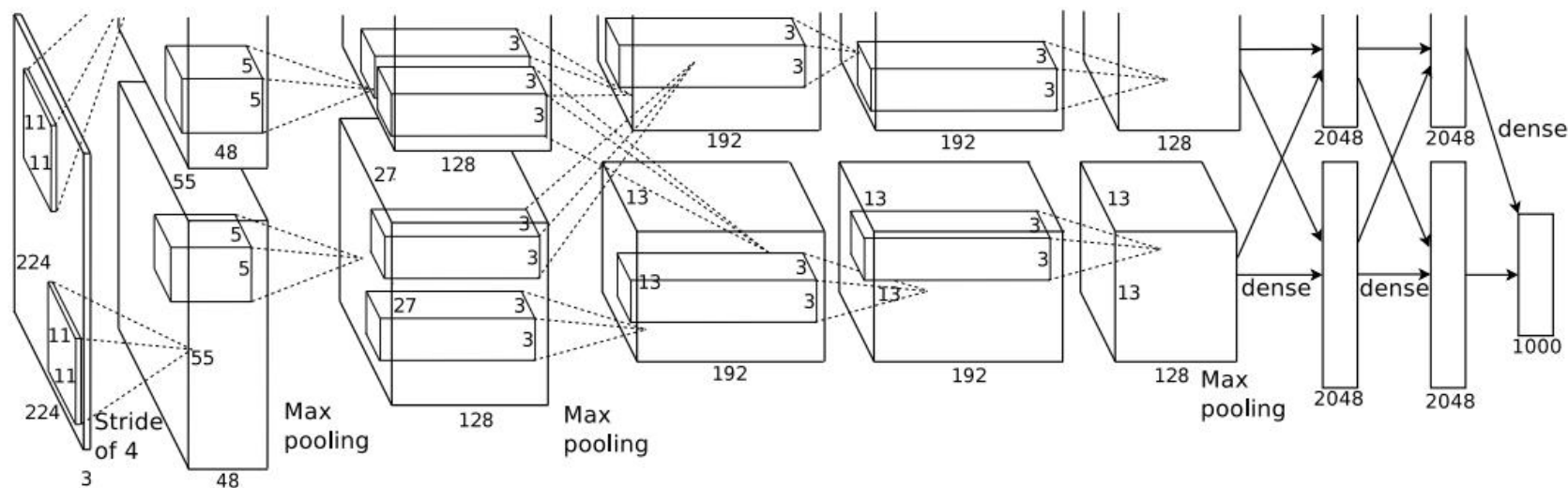


理论分析

+

实践经验

# AlexNet



- 非线性激活函数：ReLU
- 防止过拟合：Dropout
- 正则：LRN归一化层
- 其它：百万级ImageNet图像数据、分Group实现双GPU并行

# 7. 卷积神经网络CNN

---

7.1

CNN模型简介

7.2

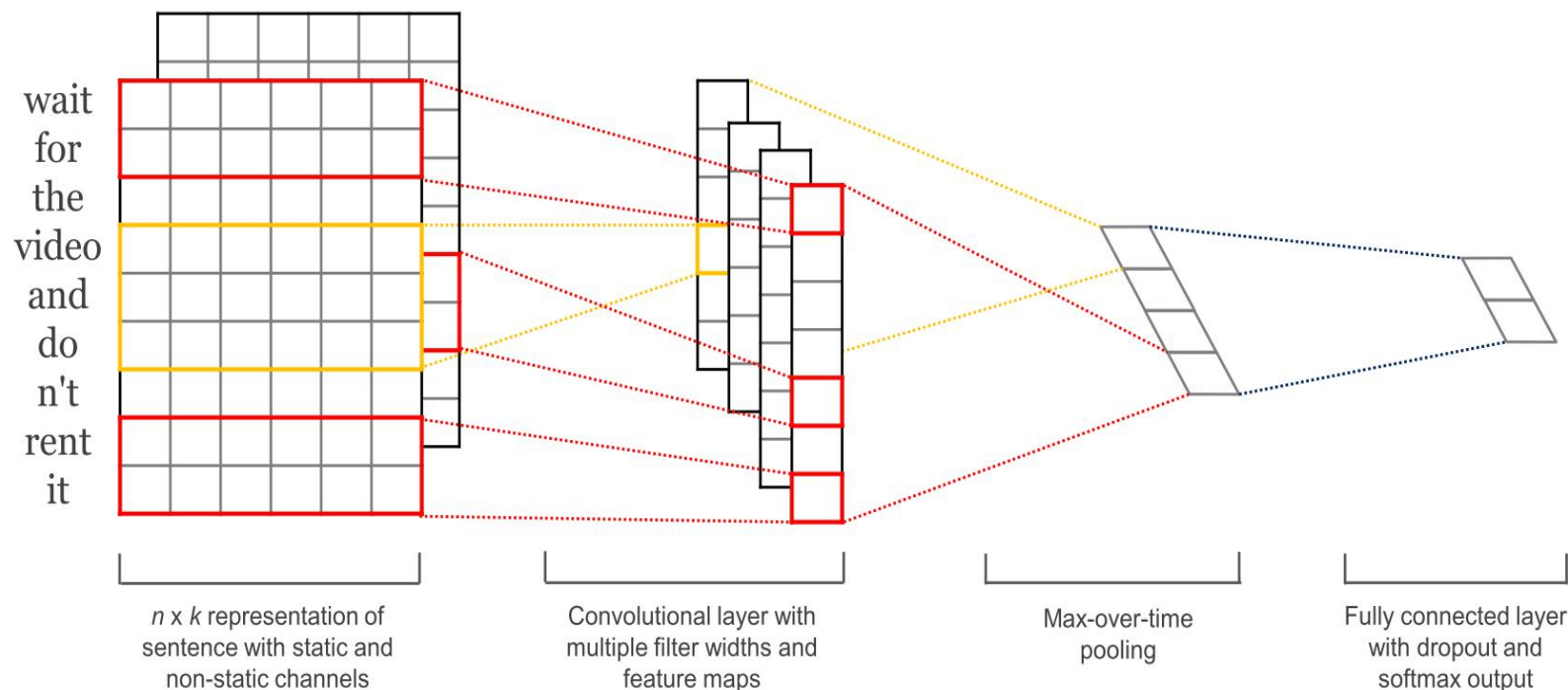
CNN的超参数

7.3

CNN for NLP



# CNN for Sentence Classification



- 输入：一句话的矩阵表示，每行是一个词向量

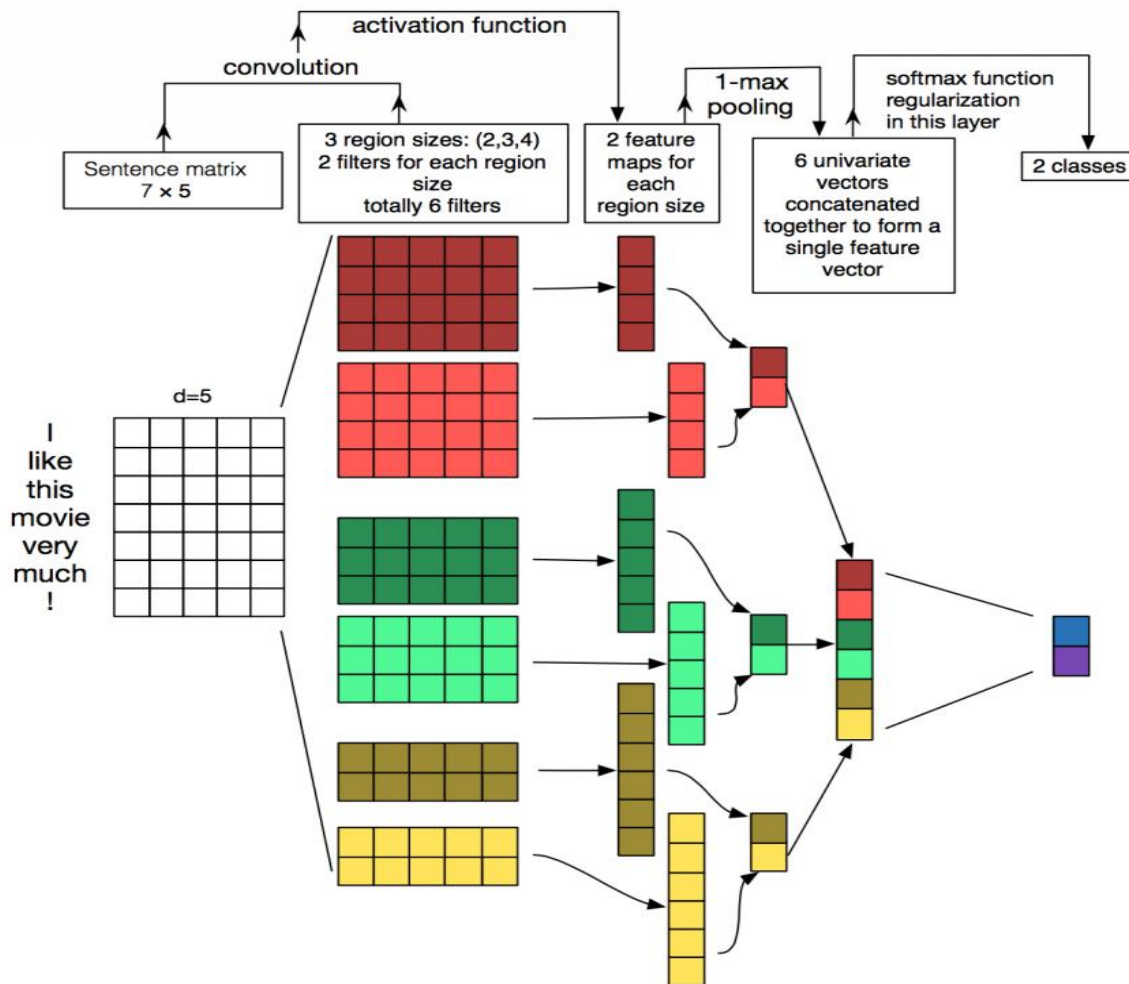
- 一次卷积：卷积核宽度等于词向量的维度

- 一次最大池化：将变长的输入转换成固定长度

- 全连接层（使用 dropout）和 softmax

# CNN for Sentence Classification

## —Hyperparameters



### 输入层

- 词向量: GloVe, word2vec, one-hot

### 卷积层

- 不同高度的卷积核 (1~10)

### 池化层

- 最大池化、平均池化
- $k$ 最大池化,  $k$ 的大小

### 激活函数

- Tanh, ReLU, Sigmoid, Softplus, Iden等7种

### 正则

- dropout和l2-norm

A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification, 2015

# CNN for Sentence Classification

## ——Hyperparameters

- 输入：GloVe、word2vec、one-hot
  - GloVe和word2vec在句子分类中的效果各有千秋
  - One-hot在文档分类中效果可以，句子分类中效果不好
- 卷积核大小
  - 不是越大越好，也不是越小越好（1~10）
  - 组合不同大小的卷积核效果更好
- 池化策略
  - 最大池化比平均池化好，1-max pooling最好
- 激活函数
  - 单层CNN 中Tanh、ReLU、Iden效果最好
- 正则
  - dropout和l2-norm效果不大，可能由于单层CNN参数不多，使用word embedding也能很好地避免过拟合

# New Ideas

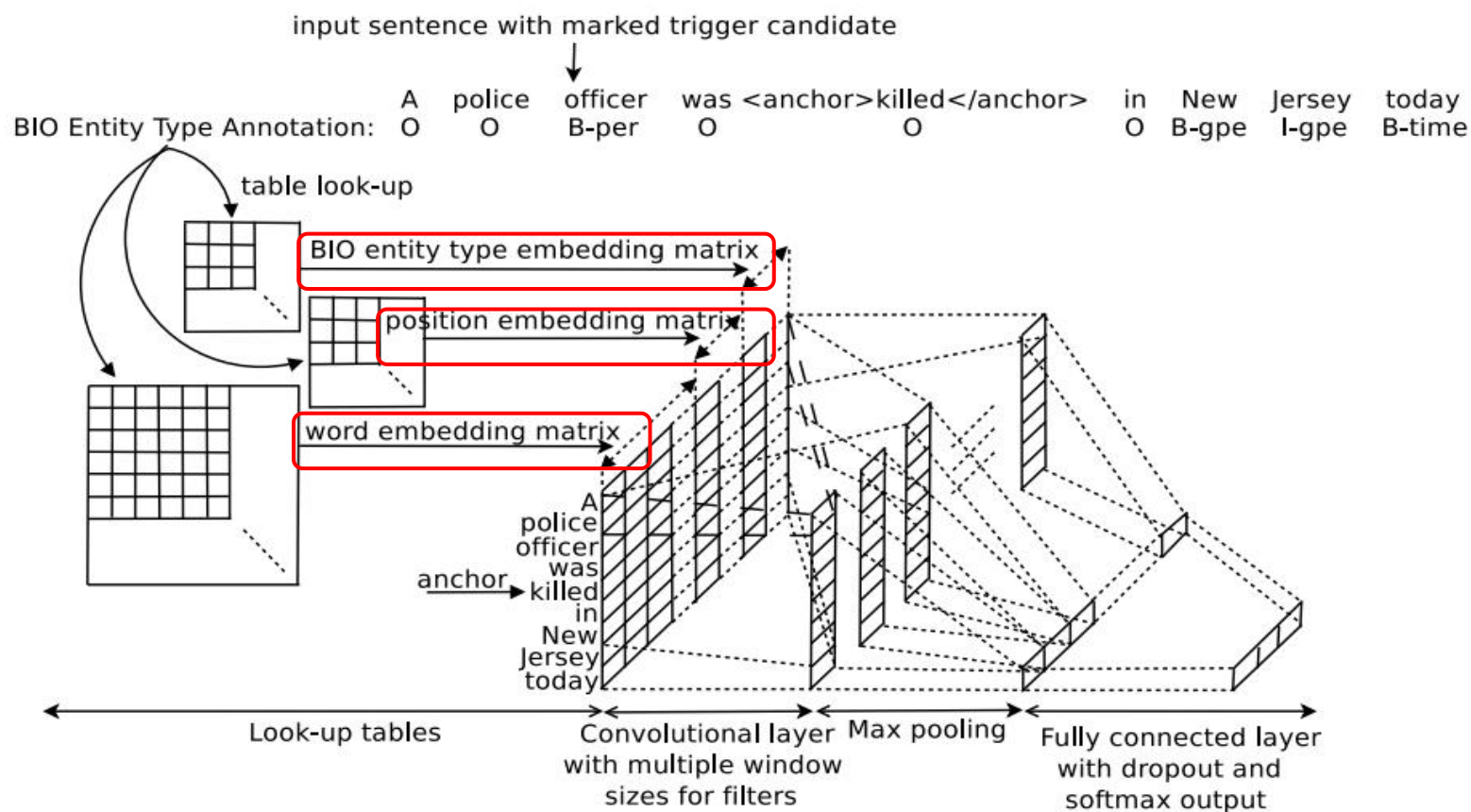
---



- 改造输入层
- 改造卷积层
- 改造池化层
- CNN模型组合

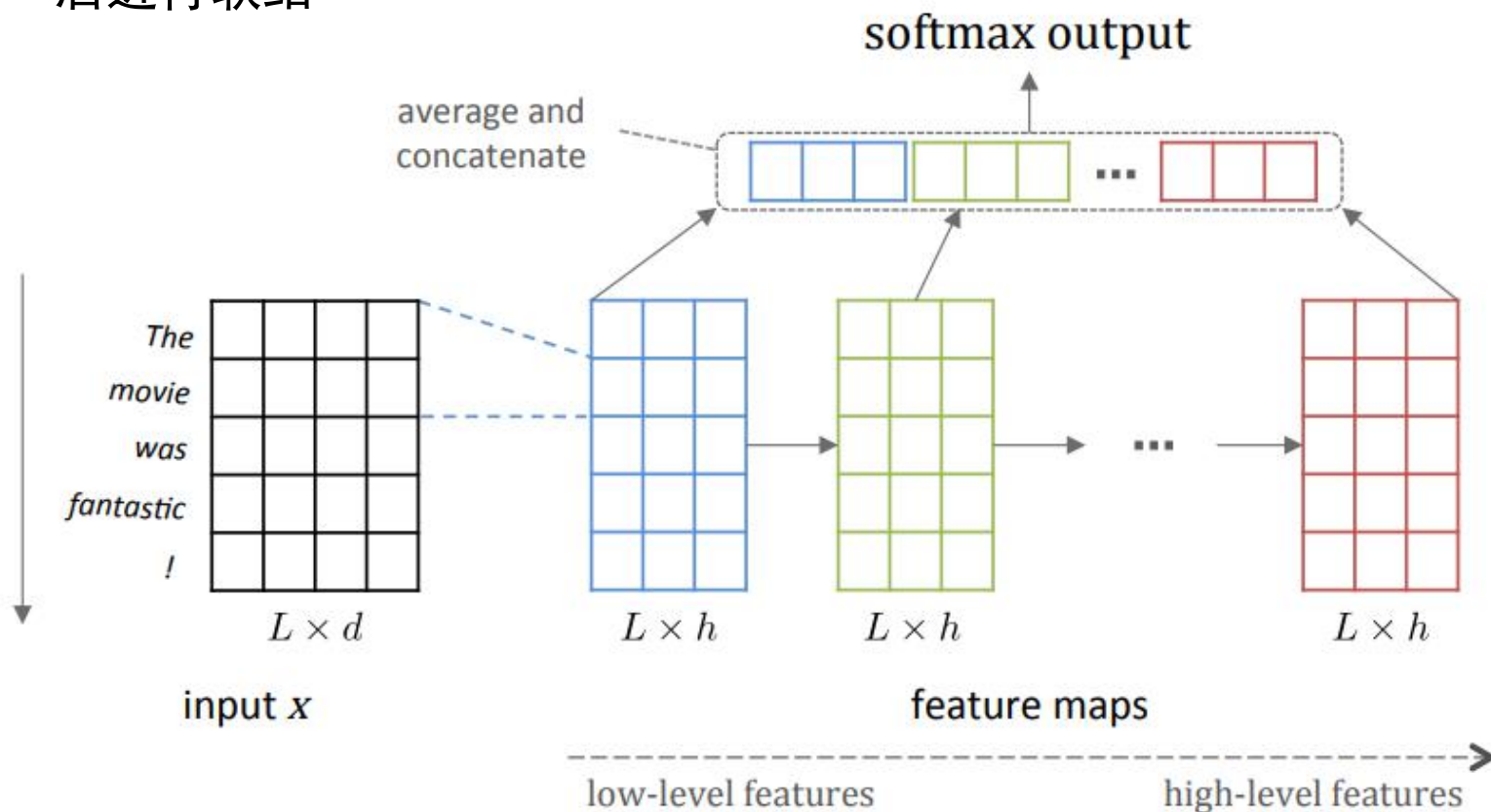
# 扩展CNN的输入

- 在事件发现中，使用三种词向量的拼接：词向量、BIO实体类型向量、位置嵌入向量



# 对卷积层进行改造

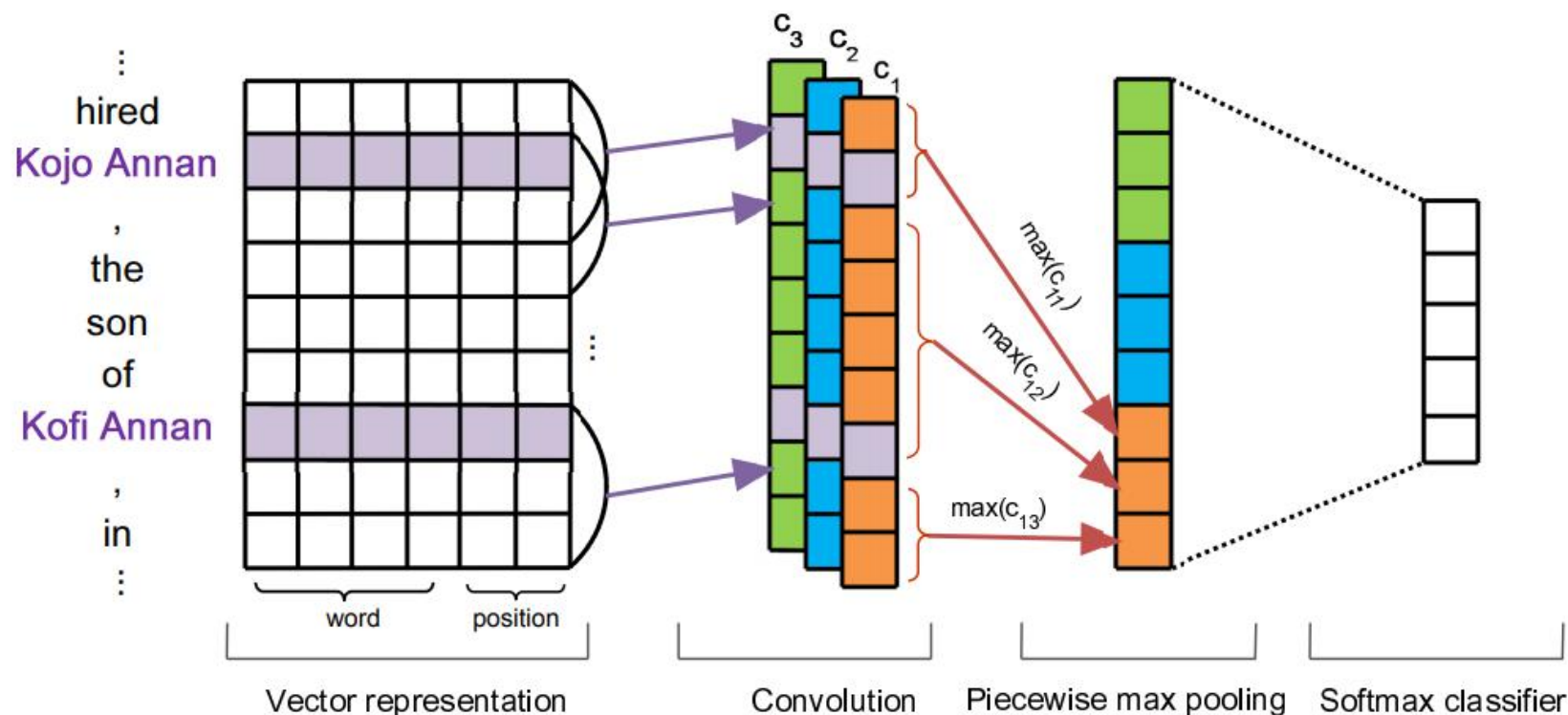
- 将卷积操作由线性操作转换成非线性、非连续操作
- 将输入矩阵映射到不同层次上的特征表示，每个层次的特征求平均后进行联结



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

# 对Pooling层进行改造

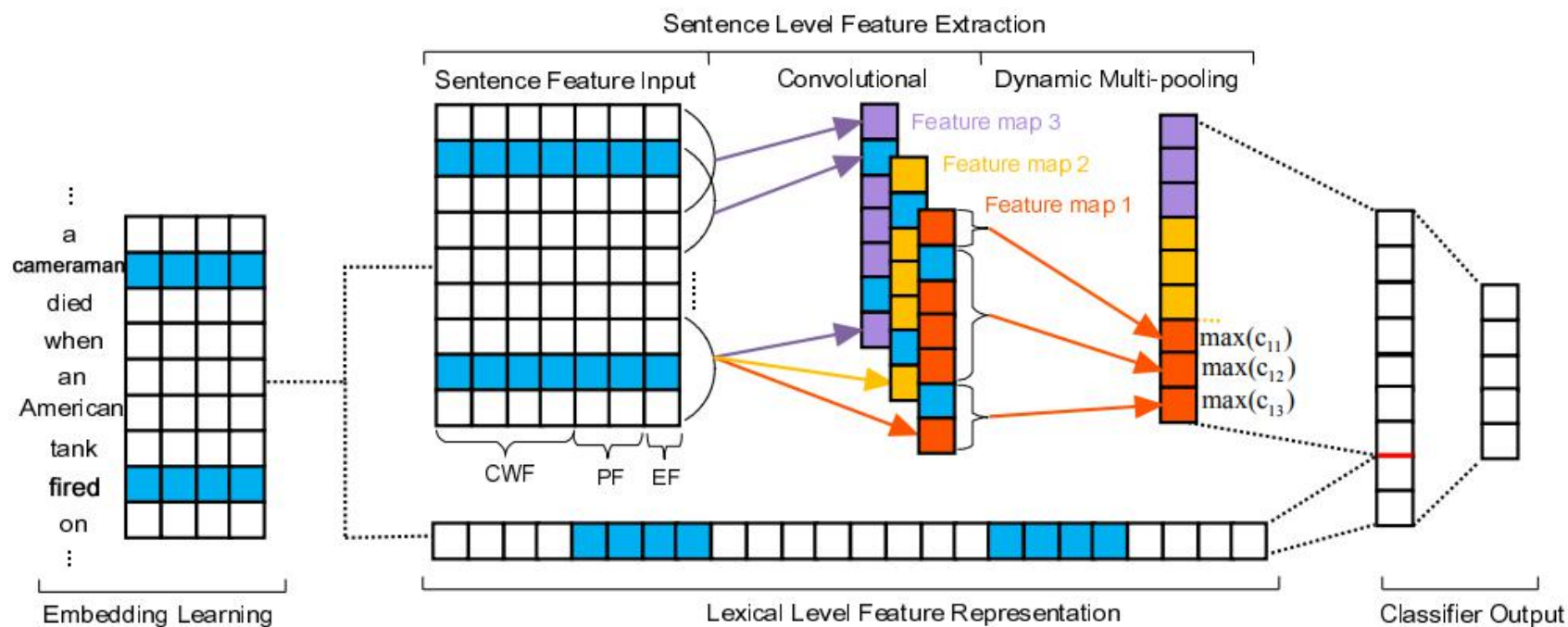
- 实体关系抽取：对输入序列中的两个实体之间的关系进行分类
- 两个实体把文本分成了三段，采用分段池化学习实体与实体之间的结构特征





# 对Pooling层进行改造

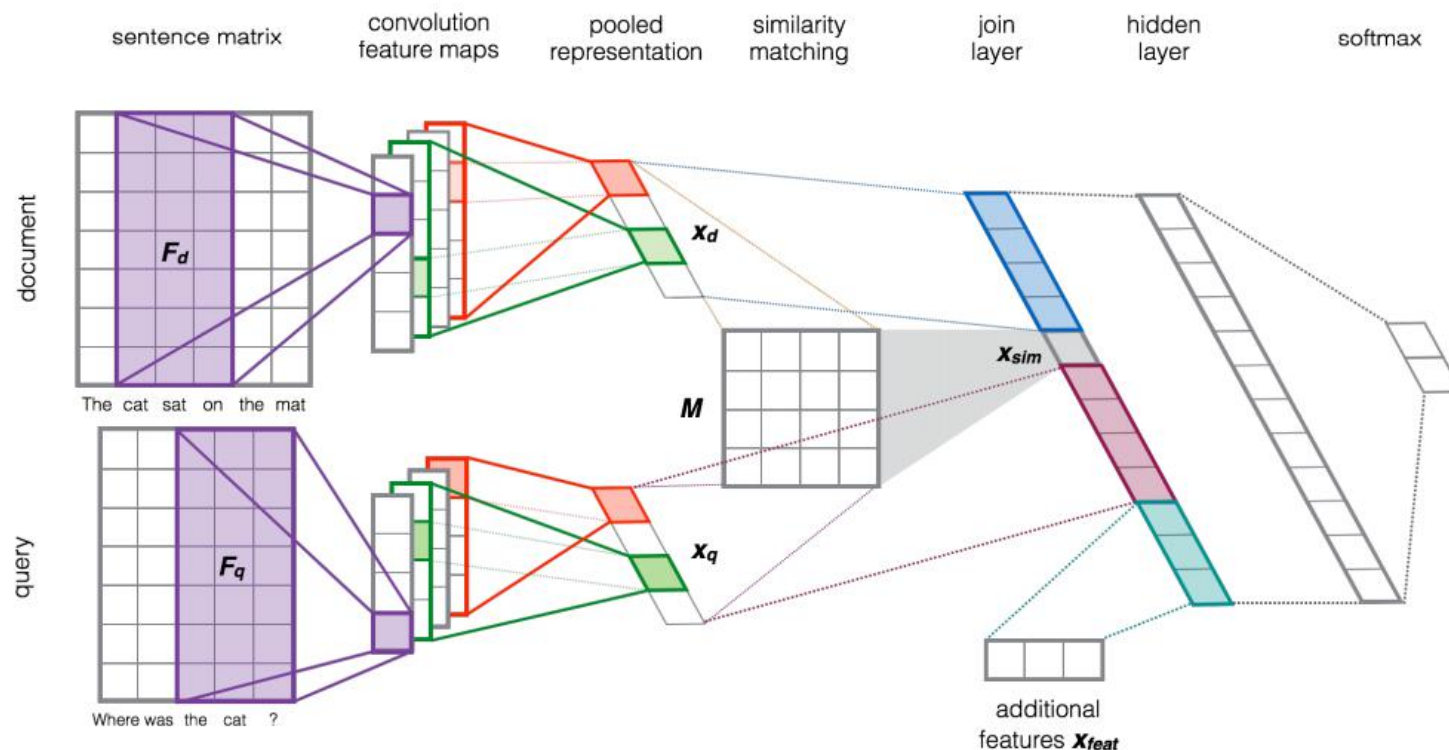
- 事件抽取：对事件中的论元关系进行分类
- 采用分段池化学习触发词与论元之间的结构特征





# CNN模型的组合

- 在短文本匹配中，对query和document使用独立的卷积和池化操作，得到它们的特征向量
- 对两个向量进行相似度计算，并将相似度值和query、document向量拼接成一个向量，输入到全连接层和softmax层



# CNN for NLP 小结

---

- 模型结构

- 大多沿用基础的CNN模型结构，包括1层卷积和1层池化
- 输入数据的不同通道采用不同的词向量表示
- 卷积核的宽度与输入词向量的维度相同

- 超参数

- 调参的作用有时大于模型本身
- 词向量、卷积核大小、池化策略、激活函数的选择都很重要

- 应用领域

- 各种分类任务：文本分析、情感分析、实体关系抽取等等
- 用于其它任务的特征提取，与其它网络模型相结合

# 参考文献

---

- Lectures

- CS231n Convolutional Neural Networks for Visual Recognition
- CS224d Lecture13: Convolutional Neural Networks (for NLP)

- Papers

- Gradient-Based Learning Applied to Document Recognition
- Convolutional Neural Networks for Sentence Classification
- Event Extraction via Dynamic Multi-Pooling Convolutional Neural Networks
- A Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification
- Event Detection and Domain Adaptation with Convolutional Neural Networks
- Molding CNNs for text: non-linear, non-consecutive convolutions
- Distant Supervision for Relation Extraction via Piecewise Convolutional Neural Networks
- Learning to Rank Short Text Pairs with Convolutional Deep Neural Networks
- Learning Mid-Level Features For Recognition

# 欢迎加入DL4NLP!



中国科学院 信息工程研究所  
INSTITUTE OF INFORMATION ENGINEERING, CAS