

## Lecture Notes: Part VIII

### Convolutional Neural Networks



CS224n 是顶级院校斯坦福出品的深度学习与自然语言处理方向专业课程，核心内容覆盖 RNN、LSTM、CNN、transformer、bert、问答、摘要、文本生成、语言模型、阅读理解等前沿内容。

#### 笔记核心词:

CNN, 卷积核, n-gram, filter, Multiple-Channels, k-max pooling, 卷积神经网络, 文本分类

课程**全部资料和信息**已整理发布, 扫描下方二维码**任意**二维码, 均可获取!!



**微信公众号** · 全套资料

回复 **CS224n**

底部**菜单栏**



**Bilibili** · 课程视频

视频**简介**

置顶**评论**



**GitHub** · 项目代码

阅读 **ReadMe**

点击**超链接**

# 1. CNNs (Convolutional Neural Networks)

## 1.1 Why CNNs?

卷积神经网络接收词向量的序列，并首先为所有子短语创建短语向量，而不仅仅是语法正确的短语(与递归神经网络一样，将在下一组笔记中讨论)。然后，CNNs 将他们分组完成手头的任务。

## 1.2 What is Convolution?

我们从一维的情况开始。考虑两个一维向量  $f$  和  $g$ ，其中  $f$  是主向量， $g$  是 filter。 $f$  和  $g$  之间的卷积，第  $n$  项处的值表示为  $(f * g)[n]$ ，它等于  $\sum_{m=-M}^M f[n-m]g[m]$ 。

右图 1 显示了二维卷积的情况。 $9 \times 9$  的绿色矩阵表示关注的主矩阵  $f$ 。 $3 \times 3$  的红色矩阵表示 filter  $g$ ，当前正在计算的卷积位于位置[2,2]。图 1 显示了第二个表中[2,2]=4 处卷积的值。你能完成第二张表格吗？

## 1.3 A Single-Layer CNN

考虑单词向量  $x_i \in R^k$  和一个  $n$  个单词的句子的单词向量串联， $x_{1:n} = x_1 \oplus x_2 \dots \oplus x_n$ 。最后，考虑卷积滤波器  $w \in R^{hk}$ ，即作用于  $h$  个单词。对于  $k=2, n=5, h=3$ ，图 2 为 NLP 的单层卷积层。在"the country of my birth"这个句子中，连续三个单词的每一个可能组合都将得到一个值。注意，滤波器  $w$  本身是一个向量，我们将有  $c_i = f(w^T x_{i:i+h-1} + b)$  来给出  $c = [c_1, c_2 \dots c_{n-h+1}] \in R^{n-h+1}$ 。最后两个时间步，即从“my”或“birth”开始，我们没有足够的字向量来与滤波器相乘(因为  $h=3$ )。如果我们需要与后两个词向量相关的卷积，一个常见的技巧是用  $h-1$  个零向量填充句子的右边，如右图 3 所示。

### 1.1 Notes info.

课件/Slides Lecture 11, P16

视频/Video Lecture 11, 8:00

GitHub · 代码 实时在线查阅文档

Bilibili · 视频 中英字幕课程视频

Stanford University X ShowMeAI

### 1.2 Notes info.

课件/Slides Lecture 11, P9

视频/Video Lecture 11, 11:00

GitHub · 代码 实时在线查阅文档

Bilibili · 视频 中英字幕课程视频

Stanford University X ShowMeAI

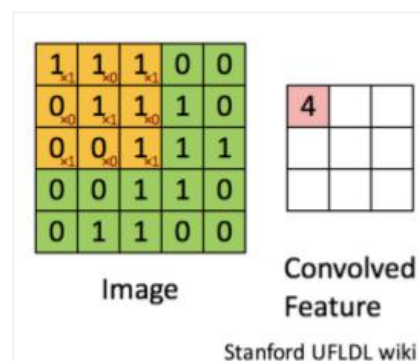


Figure 1: Convolution in the 2D case 【图 1: 2D 情况下的卷积】

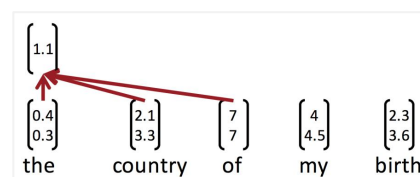


Figure 2: Single-Layer Convolution: one-step 【图 2: 单层卷积：一步】

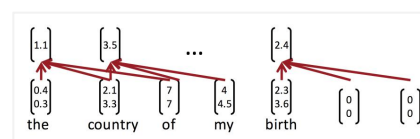


Figure 3: Single-Layer Convolution: all-steps 【图 3: 单层卷积：所有步骤】

## 1.4 Pooling

假设我们不使用补零，我们将得到最终的卷积输出， $\mathbf{c}$  有  $n - h + 1$  个数。通常，我们希望接收 CNN 的输出，并将其作为输入，输入到更深层，如前馈神经网络或 RecNN。但是，所有这都需要一个固定长度的输入，而 CNN 输出的长度依赖于句子的长度  $n$ 。解决这个问题一个聪明的方法是使用 max-pooling。CNN 的输出  $\mathbf{c} \in \mathbb{R}^{n-h+1}$  是 max-pooling 层的输入。max-pooling 的输出层  $\hat{c} = \max\{\mathbf{c}\}$ ，因此  $\hat{c} \in \mathbb{R}$ 。

我们也可以使用最小池化，因为通常我们使用 ReLU 作为非线性激活函数而 ReLU 的下界是 0。因此，一个最小池化可能会被 ReLU 覆盖，所以我们几乎总是使用最大池化而不是最小池化。

## 1.5 Multiple-Filters

在上面与图 2 相关的例子中，我们有  $h = 2$ ，这意味着我们只使用一个特定的组合方法，即使用过滤器来查看 bi-grams。我们可以使用多个 bi-grams 过滤器，因为每个过滤器将学习识别不同类型的 bi-grams。更一般地说，我们并不仅限于使用 bi-grams，还可以使用 tri-grams、quad-grams 甚至更长的过滤器。每个过滤器都有一个关联的最大池化层。因此，CNN 层的最终输出将是一个长度等于过滤器数量的向量

## 1.6 Multiple-Channels

如果我们允许梯度流入这里使用的单词向量，那么单词向量可能会随着训练而发生显著变化。这是需要的，因为它将单词向量专门用于当前特定任务(远离 GloVe 初始化)。但是，如果单词只出现在测试集中而没有出现在训练集上呢？虽然出现在训练集中的其他语义相关的单词向量将从它们的起始点显著移动，但是这些单词仍将处于它们的初始点。神经网络将专门用于已更新的输入。因此，我们在使用这些单词的句子中会表现得很差。

一种方法是维护两组单词向量，一组“静态”(没有梯度流)和一组“动态”(通过 SGD 更新)。它们最初是一样的 (GloVe 或者其他初始化)。这两个集合同时作为神经网络的输入。因此，初始化的词向量在神经网络的训练中始终起着重要的作用。在测试中给出看不见的单词可以提高正确理解的几率。

### 1.4 Notes info.

课件/Slides	Lecture 11, P20
视频/Video	Lecture 11, 33:00
GitHub · 代码	实时在线查阅文档
Bilibili · 视频	中英字幕课程视频

Stanford University X ShowMeAI

### 1.5 Notes info.

课件/Slides	Lecture 11, P24
视频/Video	Lecture -, 00:00
GitHub · 代码	实时在线查阅文档
Bilibili · 视频	中英字幕课程视频

Stanford University X ShowMeAI

### 1.6 Notes info.

课件/Slides	Lecture 11, P25
视频/Video	Lecture 11, 37:00
GitHub · 代码	实时在线查阅文档
Bilibili · 视频	中英字幕课程视频

Stanford University X ShowMeAI

有几种处理这两个 channel 的方法，最常见的是在 CNN 中使用之前对它们进行平均。另一种方法是将 CNN 过滤器的长度加倍。

## 1.7 CNN Options

### Narrow vs Wide

参见图 4。另一种问这个问题的方法是我们应该缩小还是扩大？即我们是否使用 zero-pad？

如果我们使用窄卷积，我们只在一个滤波器的所有分量都有一个匹配输入分量的位置计算卷积。在输入的开始和结束边界处显然不是这样，如图 4 中的左侧网络所示。

如果我们使用宽卷积，我们有一个输出分量对应于卷积滤波器的每个对齐。为此，我们必须在输入的开始和结束处填充  $h - 1$  个零。

在窄卷积情况下，输出长度为  $n - h + 1$ ，而在宽卷积情况下，输出长度为  $n + h - 1$

### k-max pooling

这是对最大池化层的概括。k-max 池化层不是只从它的输入中选择最大的值，而是选择  $k$  个最大的值（并且保持原有的顺序）。设置  $k = 1$  则是我们前面看到的最大池化层。

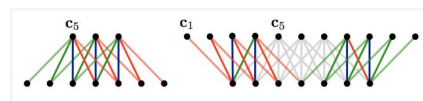


Figure 4: Narrow and Wide Convolution (from Kalchbrenner et al. (2014)) 【图 4：窄和宽卷积】

机器学习	深度学习	自然语言处理	计算机视觉	知识图谱
Machine Learning	Deep Learning	Natural Language Processing	Computer Vision	Knowledge Graphs
Stanford · CS229	Stanford · CS230	Stanford · CS224n	Stanford · CS231n	Stanford · CS520

## # 系列内容 Awesome AI Courses Notes Cheatsheets

图机器学习	深度强化学习	自动驾驶
Machine Learning with Graphs	Deep Reinforcement Learning	Deep Learning for Self-Driving Cars
Stanford · CS224W	UCBerkeley · CS285	MIT · 6.S094
...	...	...

是 ShowMeAI 资料库的分支系列，覆盖最具知名度的 TOP20+ 门 AI 课程，旨在为读者和学习者提供一整套高品质中文学习笔记和速查表。

斯坦福大学(Stanford University) Natural Language Processing with Deep Learning (CS224n) 课程，是本系列的第三门产出。

课程版本为 2019 Winter，核心深度内容(transformer、bert、问答、摘要、文本生成等)在当前(2021 年)工业界和研究界依旧是前沿的方法。最新版课程的笔记生产已在规划中，也敬请期待。

笔记内容经由深度加工整合，以 **5** 个部分构建起完整的“CS224n 内容世界”，并依托 GitHub 创建了汇总页。快扫描二维码，跳转进入吧！有任何建议和反馈，也欢迎通过下方渠道和我们联络(\*^3^\*)~

## Stanford x ShowMeAI



课程课件  
课件动态注释



课程视频  
中英字幕视频



课程笔记  
官方笔记翻译



课程作业  
作业代码解析



课程项目  
综合项目参考



微信公众号

扫码回复“CS224n”，下载最新全套资料  
扫码回复“添砖加瓦”，成为AI内容创作者