



北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

深度学习经典方法

邬成浩 李恩 徐园 刘嘉 杜建成



目录 | CONTENTS

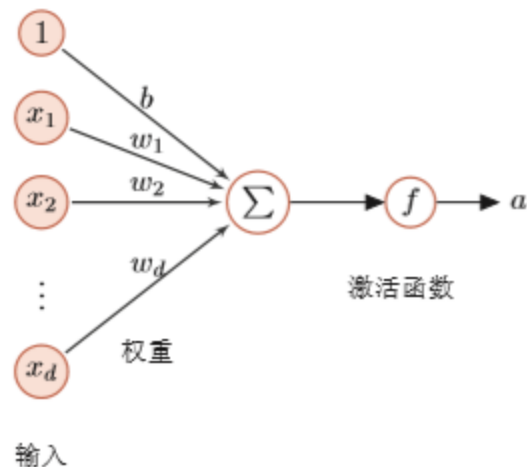
- 1 神经网络基础
- 2 卷积神经网络
- 3 循环神经网络
- 4 RNN 案例
- 5 CNN 案例



1

神经网络基础

人工神经元，简称神经元，是构成神经网络的基本单元，其主要是模拟生物神经元的结构和特性，接受一组输入信号并产生输出。对 n 个输入信号，通过带权重的连接（connection）进行传递，将总的输入与阈值进行比较，通过“激活函数”处理产生输出。



为什么引入激励函数？

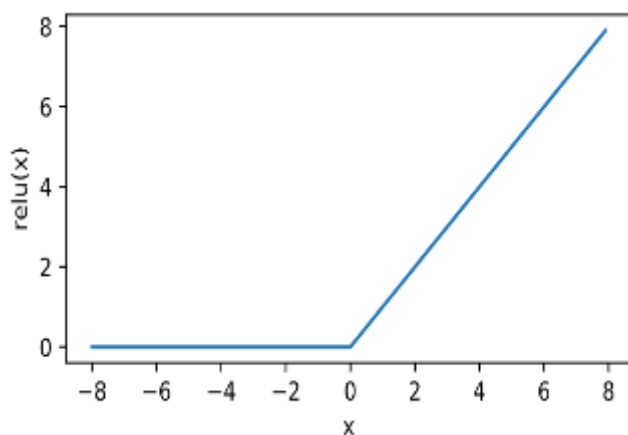
不用激励函数（其实相当于激励函数是 $f(x) = x$ ），即每一层输出都是上层输入的线性函数，神经网络有多少层，输出都是输入的线性组合，相当于没有隐藏层效果，这种情况就是最原始的感知机。

性质：

- 1.连续可导的非线性函数
- 2.简单，便于网络计算
- 3.函数值域在一个合适区间

修正线性单元 ReLU

$$\text{ReLU}(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

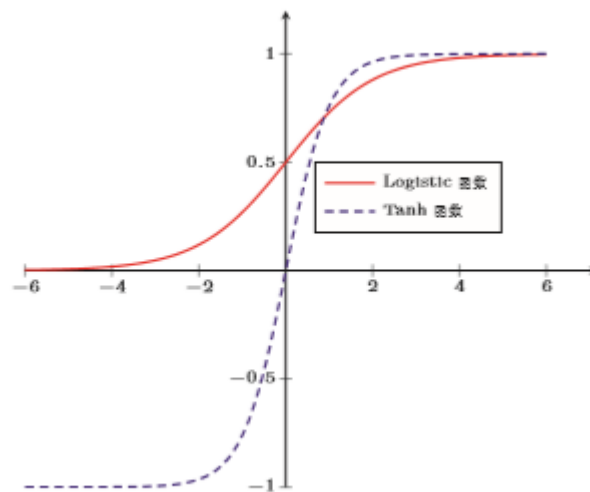


Sigmoid函数计算量大，容易出现梯度消失，
relu函数可缓解过拟合

Sigmoid型函数

logistic函数: $\sigma = \frac{1}{1 + \exp(-x)}$

Tanh函数: $\tanh(x) = \frac{\exp(x) - \exp(-x)}{\exp(x) + \exp(-x)}$





2

卷积神经网络

背景：卷积神经网络是近年来深度学习在计算机视觉领域取得突破性成果的基石。

定义：卷积神经网络(cnn)是含有卷积层(convolutional layer)神经网络。以二维为例，它有高和宽两个维度，常用来处理图像数据。

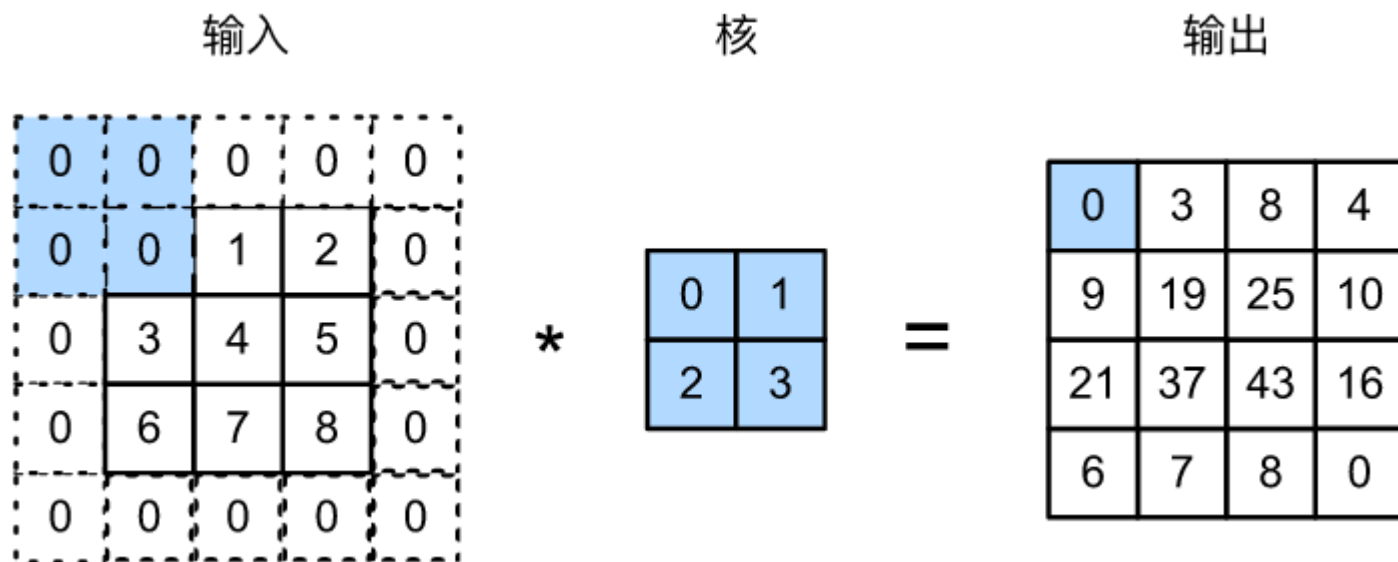
互相关运算

在二维卷积层中，一个二维输入数组和一个二维卷积核（kernel）数组通过互相关运算输出一个二维数组

输入		核		输出																	
<table border="1"><tr><td>0</td><td>1</td><td>2</td></tr><tr><td>3</td><td>4</td><td>5</td></tr><tr><td>6</td><td>7</td><td>8</td></tr></table>	0	1	2	3	4	5	6	7	8	*	<table border="1"><tr><td>0</td><td>1</td></tr><tr><td>2</td><td>3</td></tr></table>	0	1	2	3	=	<table border="1"><tr><td>19</td><td>25</td></tr><tr><td>37</td><td>43</td></tr></table>	19	25	37	43
0	1	2																			
3	4	5																			
6	7	8																			
0	1																				
2	3																				
19	25																				
37	43																				

假设输入形状是 $n_h \times n_w$ ，卷积核窗口形状是 $k_h \times k_w$ ，输出形状将会是：
 $(n_h - k_h + 1) * (n_w - k_w + 1)$

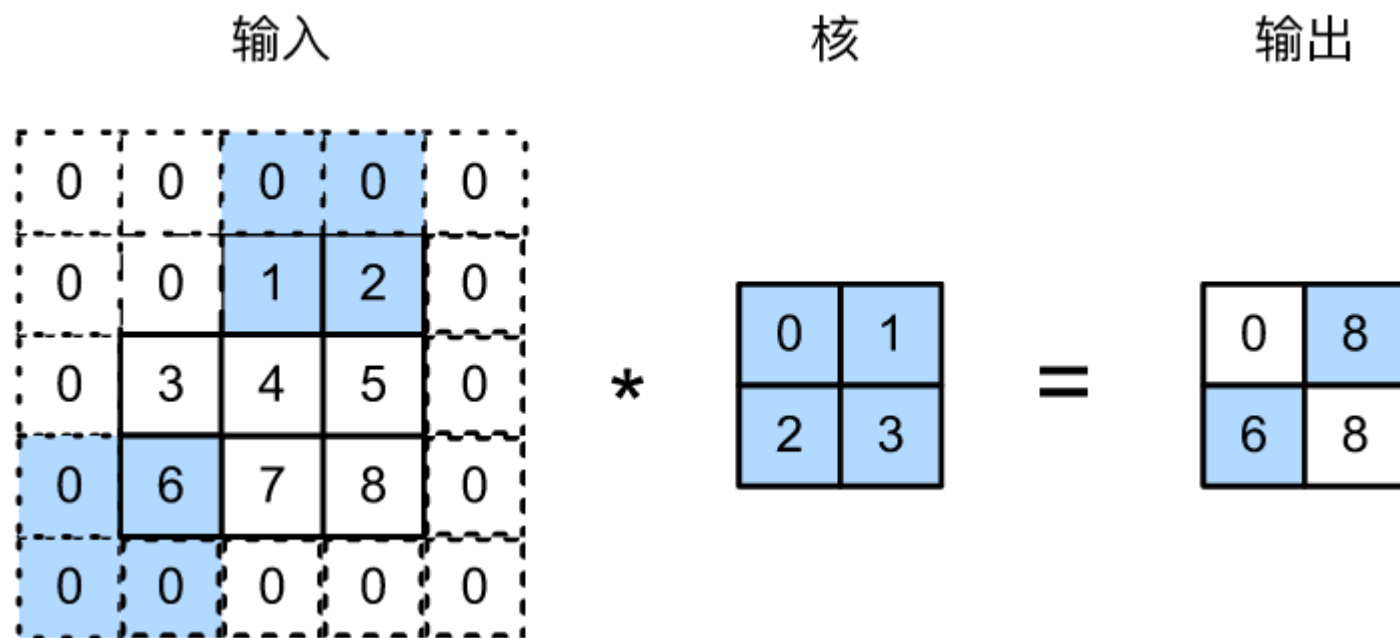
填充:填充 (padding) 是指在输入高和宽的两侧填充元素 (通常是0元素)



Why:

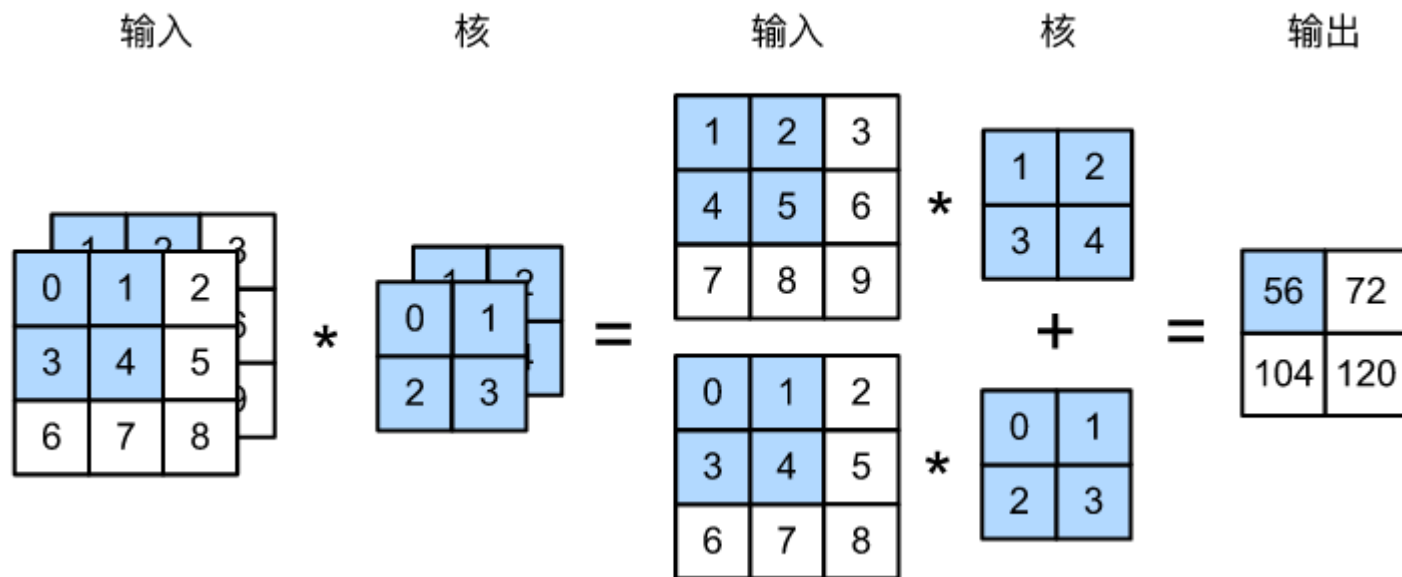
- 保持边界信息
- 图片补齐
- 输入输出维度一样

步幅：卷积窗口从输入数组的最左上方开始，按从左往右、从上往下的顺序，依次在输入数组上滑动。我们将每次滑动的行数 and 列数称为步幅 (stride)



高上步幅为3、在宽上步幅为2的二维互相关运算

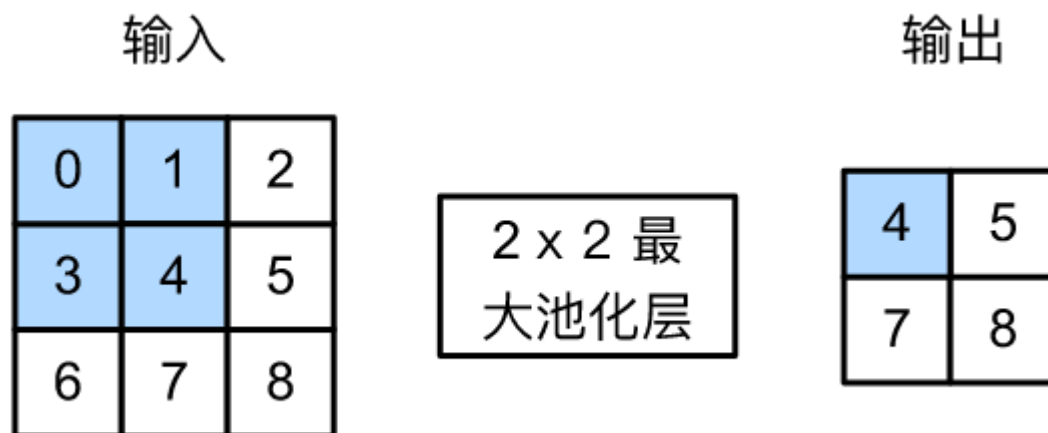
多输入通道：彩色图像在高和宽2个维度外还有RGB（红、绿、蓝）3个颜色通道。假设彩色图像的高和宽分别是h和w（像素），那么它可以表示为一个 $3 \times h \times w$ 的多维数组。我们将大小为3的这一维称为通道（channel）维。



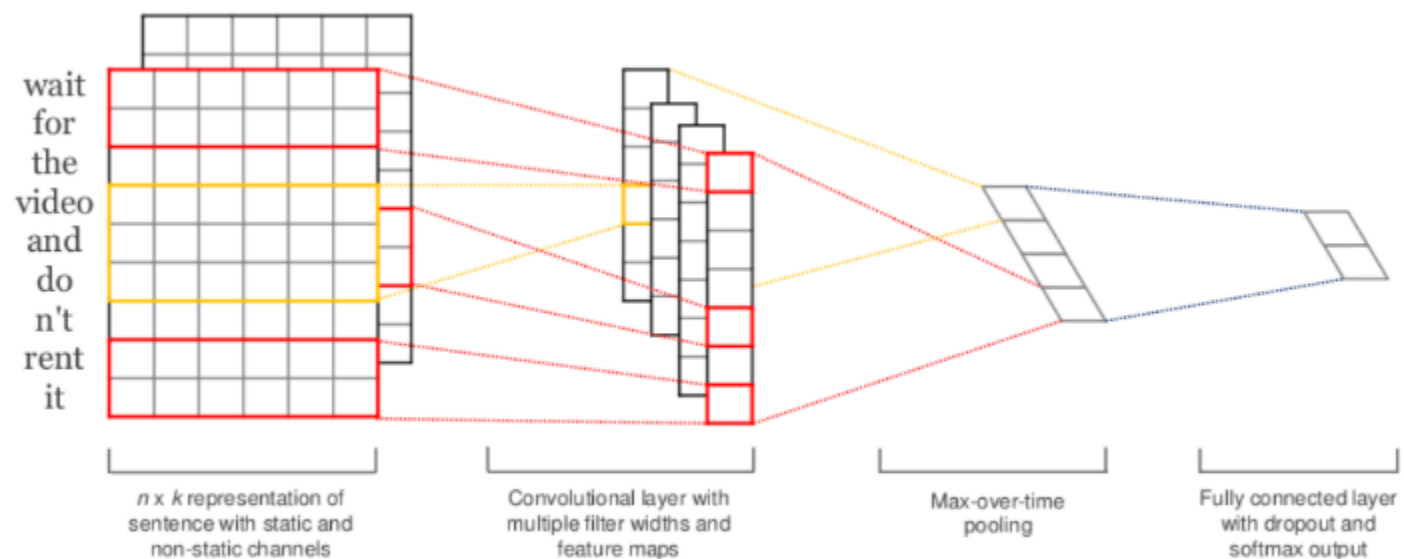
当输入数据含多个通道时，我们需要构造一个与输入数据的通道数相同的卷积核，从而能够与含多通道的输入数据做互相关运算。

池化层 (pooling)

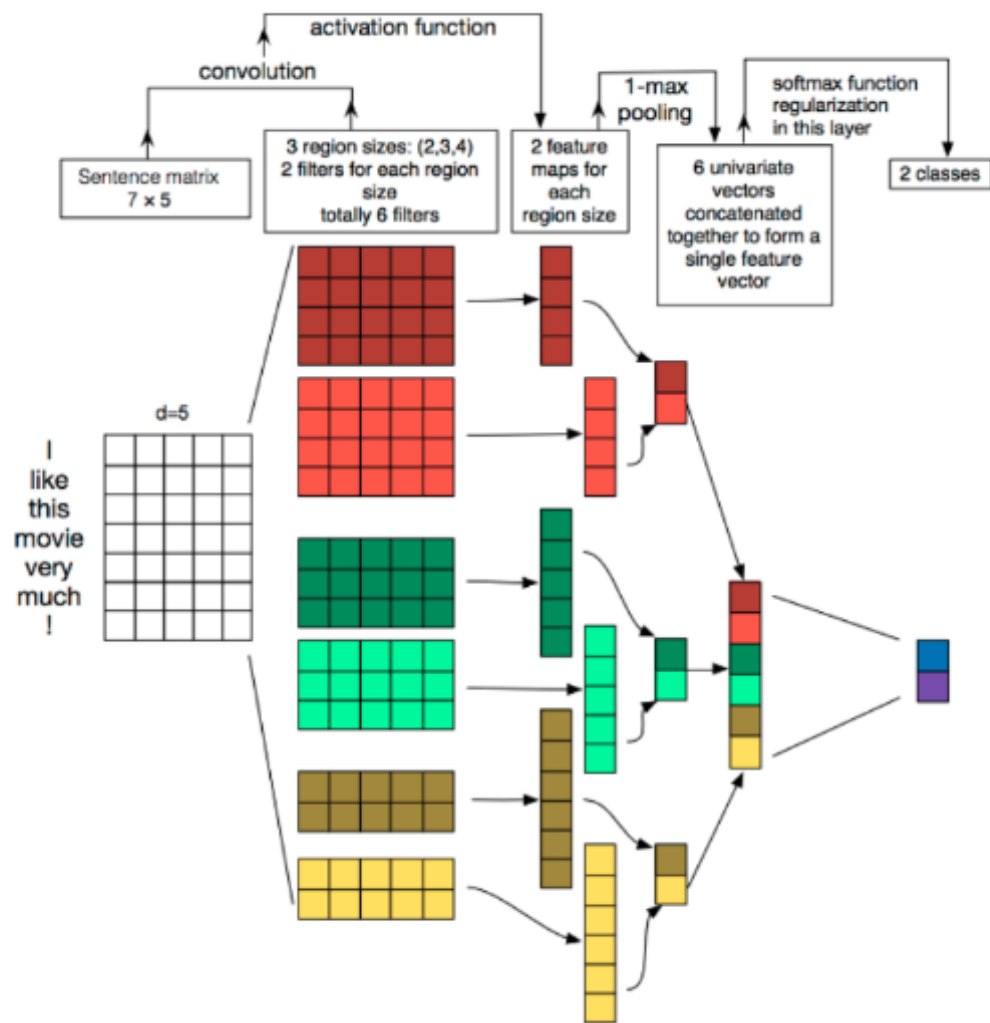
同卷积层一样，池化层每次对输入数据的一个固定形状窗口（又称池化窗口）中的元素计算输出。不同于卷积层里计算输入和核的互相关性，池化层直接计算池化窗口内元素的最大值或者平均值。池化层的一个主要作用是缓解卷积层对位置的过度敏感性。



背景：通常，卷积神经网络是用来处理二维图像数据。在之前的语言模型和文本分类任务中，文本数据看作是只有一个维度的时间序列，并很自然地使用循环神经网络来表征这样的数据。其实，可以将文本当作一维图像，从而可以用一维卷积神经网络来捕捉临近词之间的关联。2014年，Yoon Kim将卷积神经网络cnn应用到文本分类任务，利用多个不同size的kernel来提取句子中的关键信息,从而能够更好地捕捉局部相关性。



(2014 EMNLP) Convolutional Neural Networks for Sentence Classification



•**Embedding**: 第一层是图中最左边的7乘5的句子矩阵，每行是词向量，维度=5，这个可以类比为图像中的原始像素点。

•**Convolution**: 然后经过 $\text{kernel_sizes}=(2,3,4)$ 的一维卷积层，每个 kernel_size 有两个输出 channel。

•**MaxPooling**: 第三层是一个1-max pooling层，这样不同长度句子经过pooling层之后都能变成定长的表示。

•**FullConnection and Softmax**: 最后接一层全连接的 softmax 层，输出每个类别的概率。



3

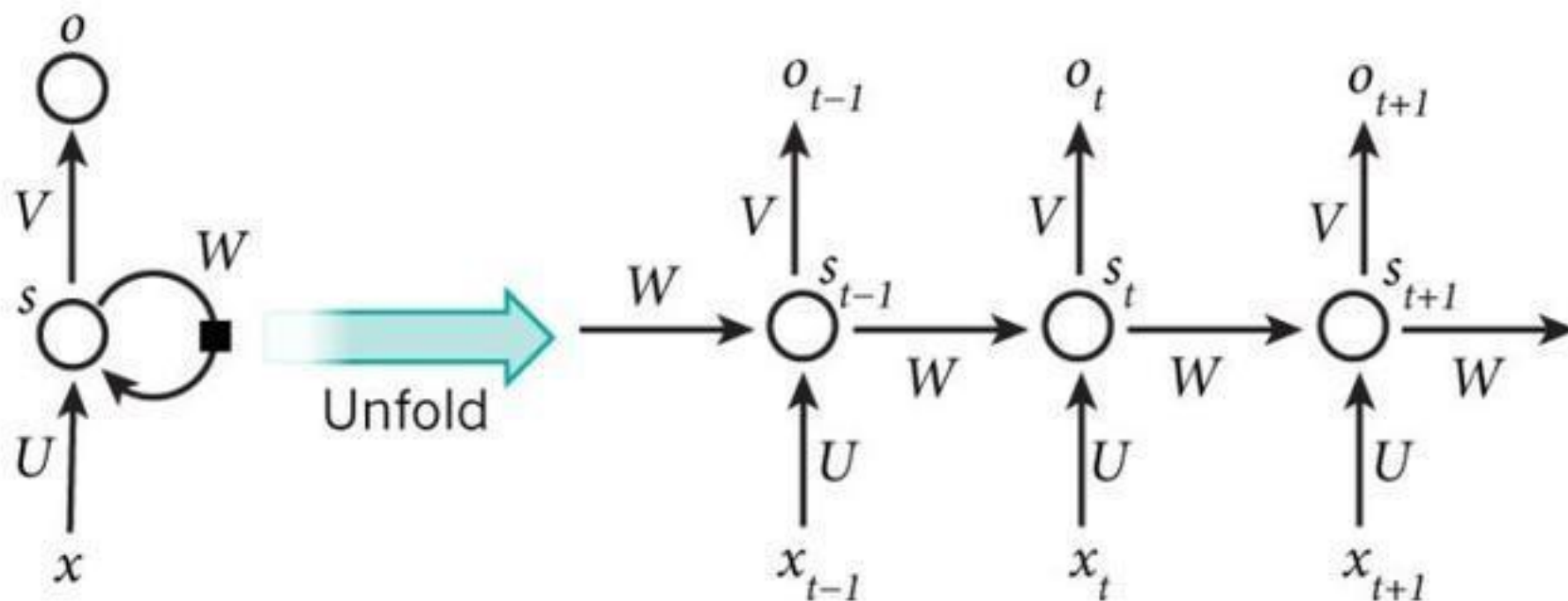
循环神经网络

简介：循环神经网络（Recurrent Neural Network, RNN）是一类处理序列数据的神经网络。

应用：自然语言处理（语音识别，机器翻译等）、计算机视觉、时间序列预测。

特点：链式连接。权值共享。

RNN的基本结构

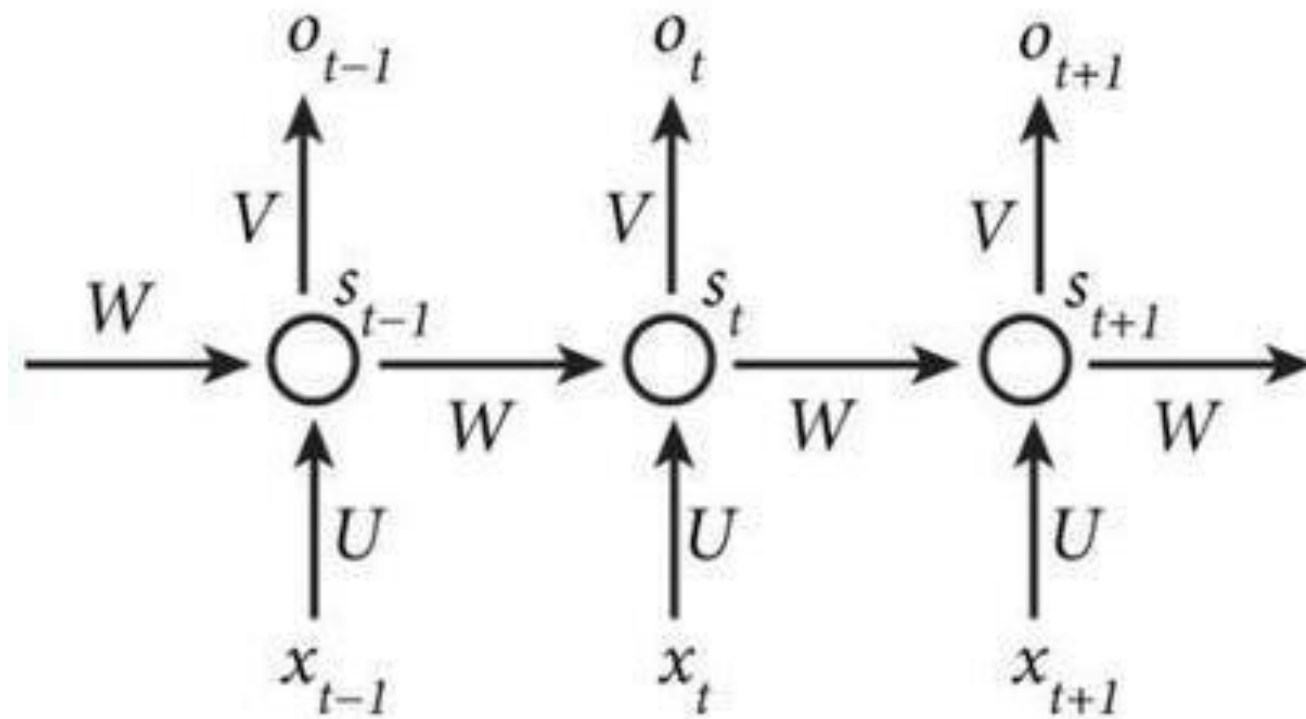


RNN的前向传播

★简单循环网络 (Simple Recurrent Network, SRN)

$$s_t = \sigma(Ws_{t-1} + Ux_t + b)$$

$$o_t = \varphi(Vs_t + c)$$



SRN存在梯度消失问题：

反向传播计算梯度时使用链式法则。

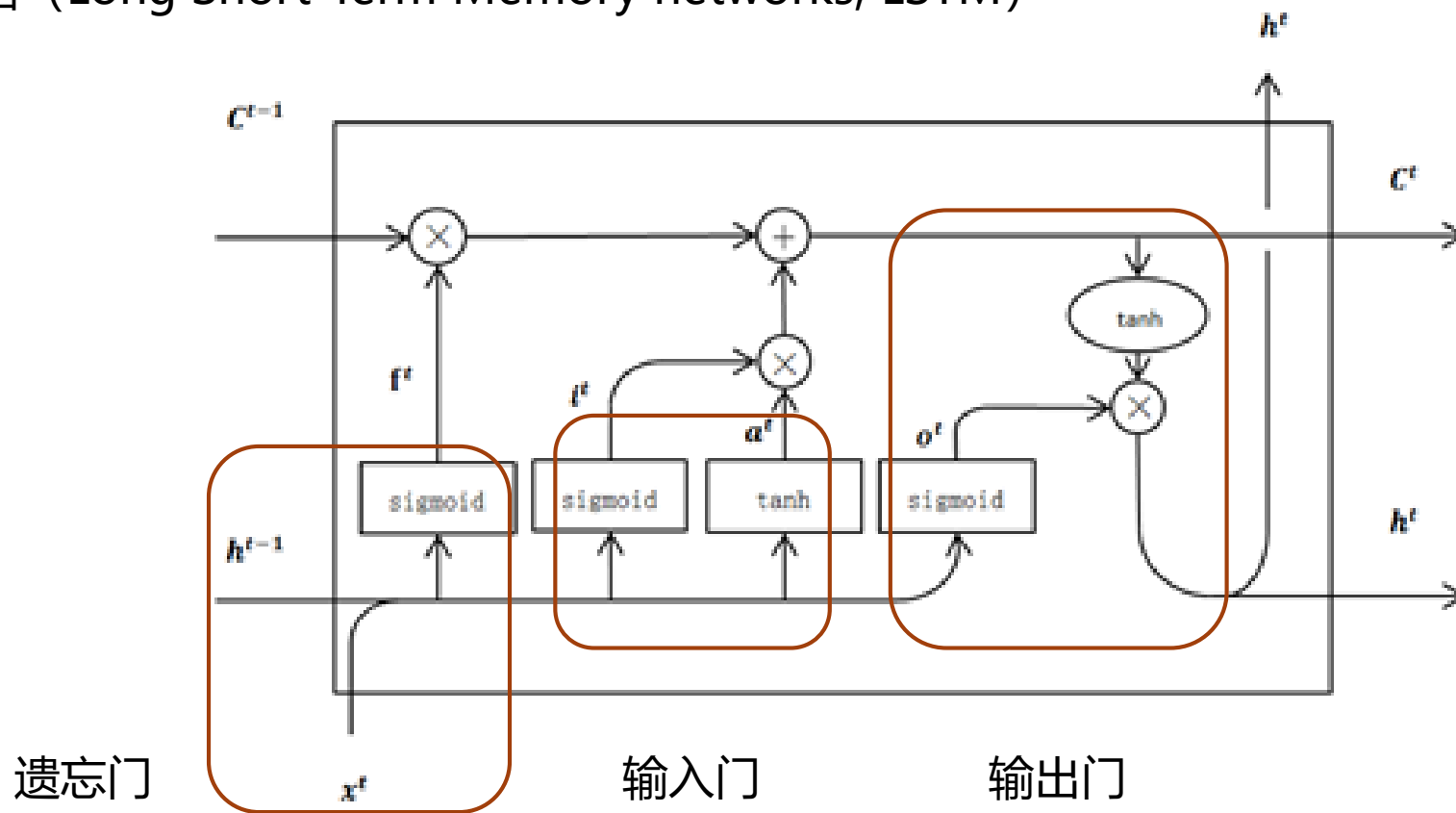
当序列过长时，累乘小数导致前面的单元梯度很小，训练速率变慢甚至无法训练。

解决思想：

引入门控单元，赋予RNN控制其内部信息积累的能力。

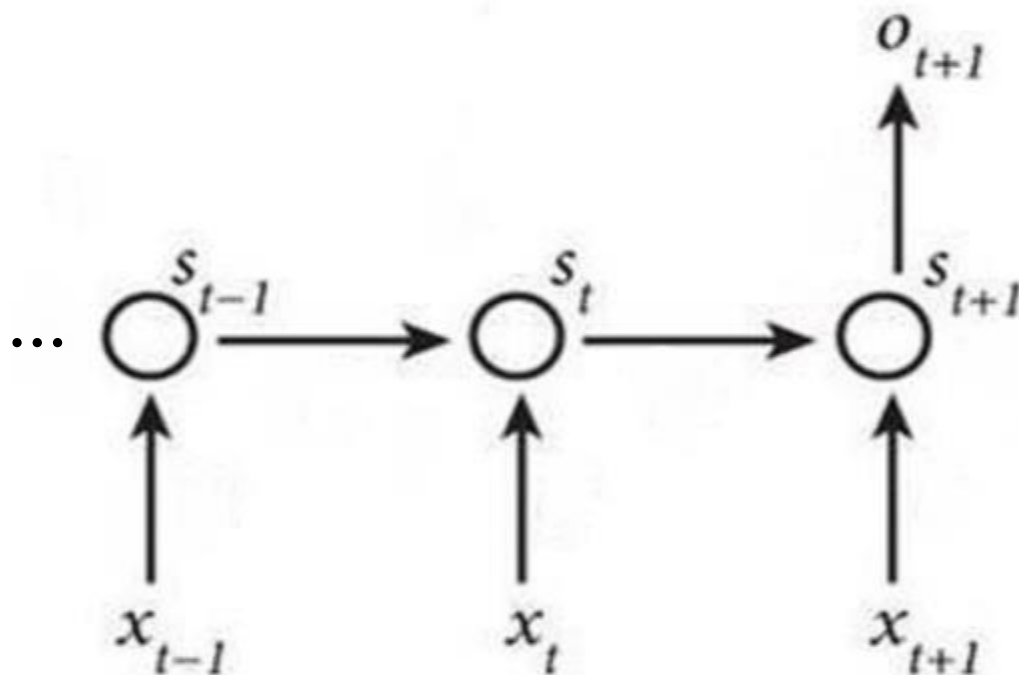
使网络既能掌握长距离依赖，又能选择性地遗忘信息防止过载。

★长短期记忆网络 (Long Short-Term Memory networks, LSTM)



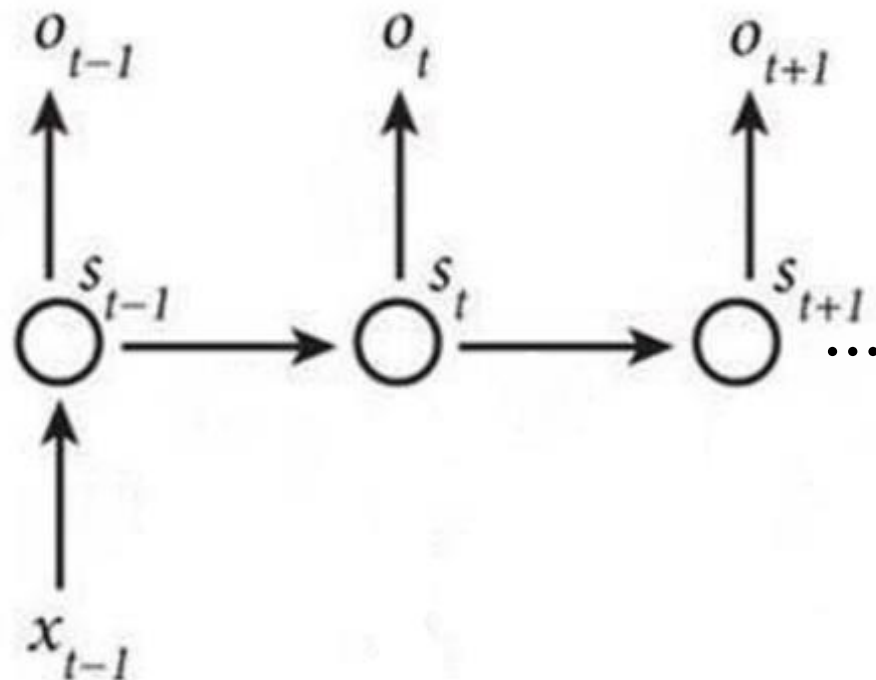
RNN的变体

应用举例：文本分类



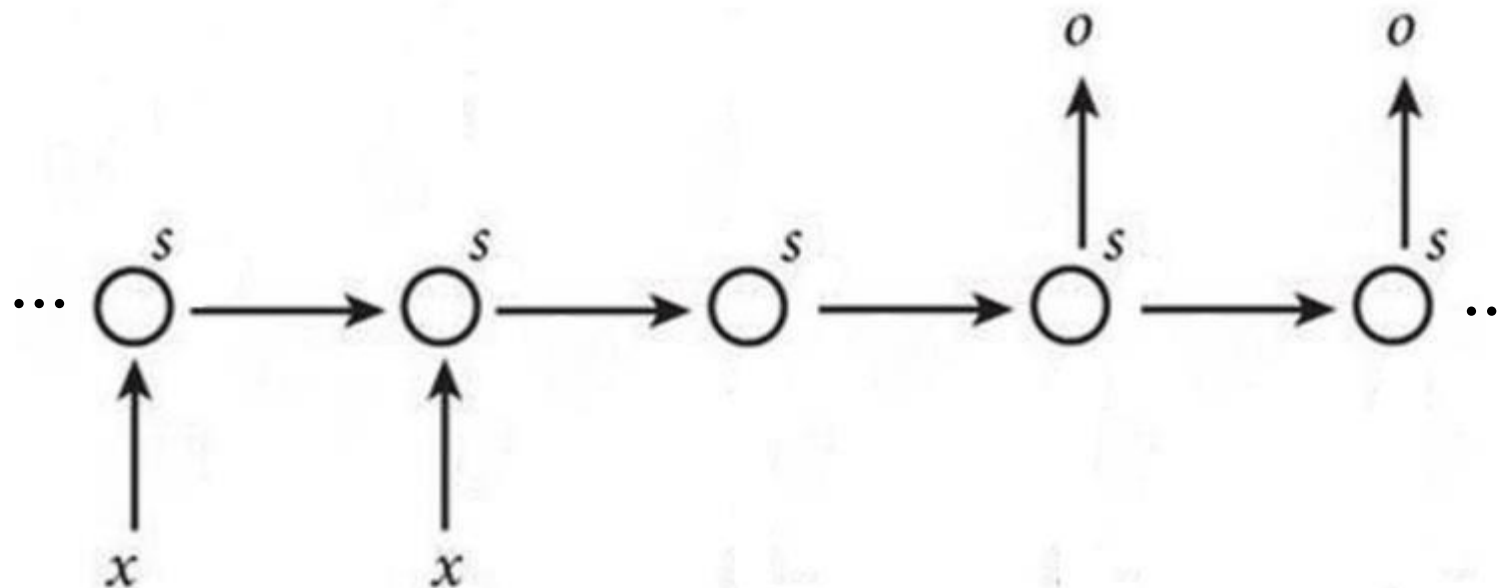
RNN的变体

应用举例：文本生成



RNN的变体

应用举例：机器翻译





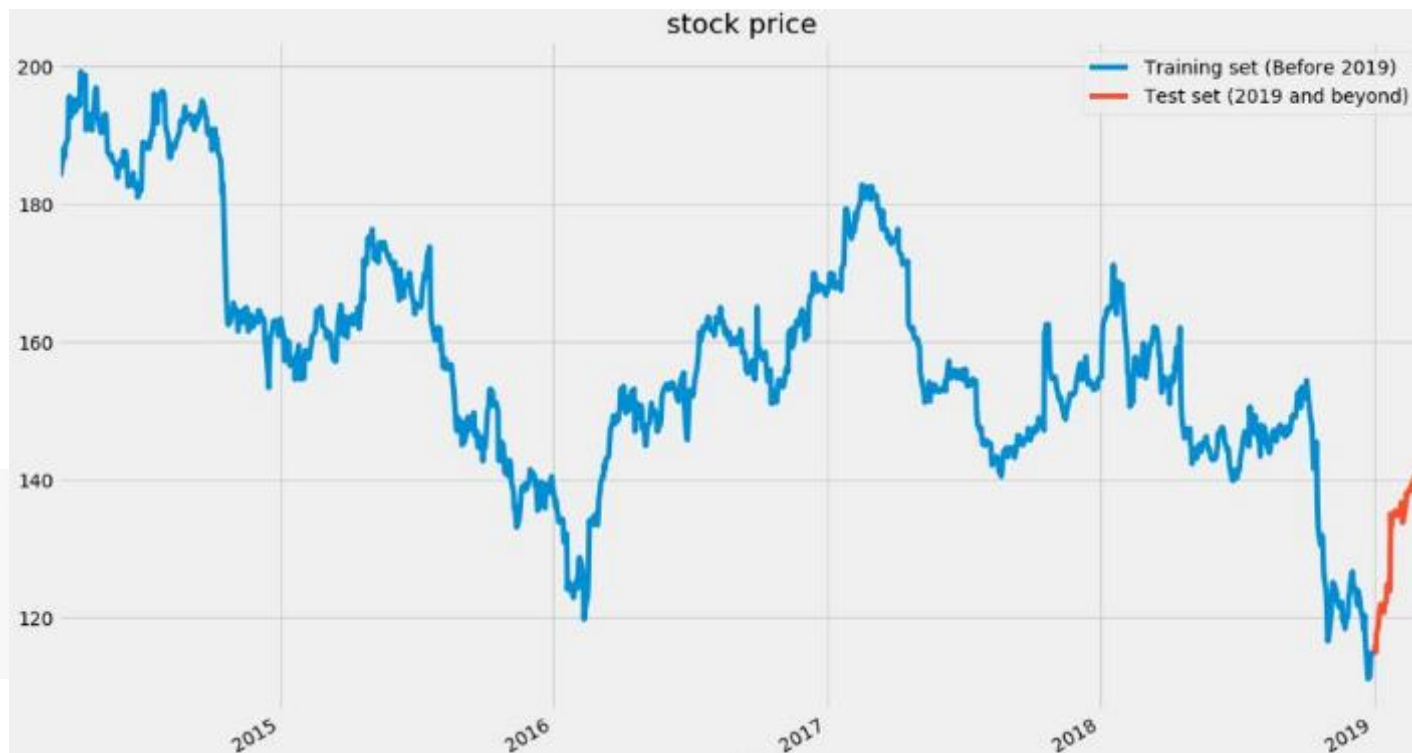
4 RNN案例

数据集

2014.3-2019.3

IBM股票当日最高价

```
X_train = []  
y_train = []  
for i in range(60, 1211):  
    X_train.append(training_set_scaled[i-60:i, 0])  
    y_train.append(training_set_scaled[i, 0])
```



网络构建与训练

```
regressor = Sequential() # 序贯模型
# 四层LSTM, 一层输出
regressor.add(LSTM(units=50, return_sequences=True, input_shape=(X_train.shape[1], X_train.shape[2])))
regressor.add(Dropout(0.1))
regressor.add(LSTM(units=50, return_sequences=True))
regressor.add(Dropout(0.1))
regressor.add(LSTM(units=50, return_sequences=True))
regressor.add(Dropout(0.1))
regressor.add(LSTM(units=50))
regressor.add(Dropout(0.1))

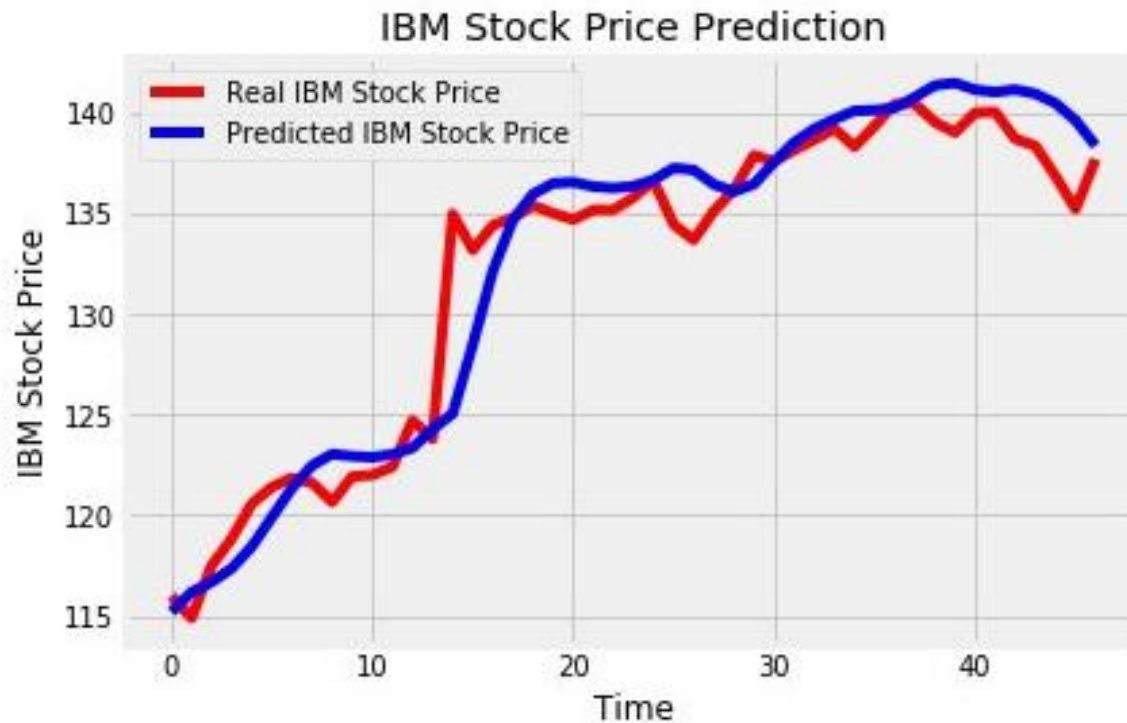
regressor.add(Dense(units=1))

# 指定优化器和损失函数
regressor.compile(optimizer='rmsprop', loss='mean_squared_error')
# 训练
regressor.fit(X_train, y_train, validation_split=0.25, epochs=50, batch_size=32)
```

```
Epoch 1/50
1151/1151 [=====] - 8s 7ms/step - loss: 0.0389
Epoch 2/50
1151/1151 [=====] - 4s 4ms/step - loss: 0.0163
Epoch 3/50
1151/1151 [=====] - 4s 4ms/step - loss: 0.0143
Epoch 4/50
1151/1151 [=====] - 4s 4ms/step - loss: 0.0128
Epoch 5/50
1151/1151 [=====] - 4s 4ms/step - loss: 0.0107
Epoch 6/50
1151/1151 [=====] - 4s 4ms/step - loss: 0.0090
Epoch 7/50
1151/1151 [=====] - 5s 4ms/step - loss: 0.0099
Epoch 8/50
1151/1151 [=====] - 5s 4ms/step - loss: 0.0080
Epoch 9/50
1151/1151 [=====] - 5s 4ms/step - loss: 0.0084
Epoch 10/50
1151/1151 [=====] - 5s 4ms/step - loss: 0.0080
Epoch 11/50
1151/1151 [=====] - 5s 4ms/step - loss: 0.0064
Epoch 12/50
1151/1151 [=====] - 5s 4ms/step - loss: 0.0068
Epoch 13/50
1151/1151 [=====] - 5s 4ms/step - loss: 0.0062
Epoch 14/50
1151/1151 [=====] - 5s 4ms/step - loss: 0.0060
Epoch 15/50
1151/1151 [=====] - 5s 4ms/step - loss: 0.0056
```

结果

测试集上的均方根误差：2.28





5

CNN案例

数据集

IMDB数据集

- 50000条偏向明显的评论
 - 25000 训练集
 - 25000 测试集
- 被标记为正面/负面 两种评价
 - Label:
 - Positive
 - Negative

	type	review	label	file
0	test	Once again Mr. Costner has dragged out a movie for far longer than necessary. Aside from the terrific sea rescue sequences, of which there are very few I just did not care about any of t	neg	0_2.txt
1	test	This is an example of why the majority of action films are the same. Generic and boring, there's really nothing worth watching here. A complete waste of the then barely-tapped talents of	neg	10000_4.txt
2	test	First of all I hate those moronic rappers, who couldnt act if they had a gun pressed against their foreheads. All they do is curse and shoot each other and acting like cliché version of	neg	10001_1.txt
3	test	Not even the Beatles could write songs everyone liked, and although Walter Hill is no mop-top he is second to none when it comes to thought provoking action movies. The nineties com	neg	10002_3.txt
4	test	Brass pictures (movies is not a fitting word for them) really are somewhat brassy. Their alluring visual qualities are reminiscent of expensive high class TV commercials. But unfortunately	neg	10003_3.txt
5	test	A funny thing happened to me while watching "Mosquito": on the one hand, the hero is a deaf-mute and the director is totally unable to make us understand why he does what he does	neg	10004_2.txt
6	test	This German horror film has to be one of the weirdest I have seen. I was not aware of any connection between child abuse and vampirism, but this is supposed based upon a	neg	10005_2.txt
7	test	Being a long-time fan of Japanese film, I expected more than this. I can't really be bothered to write too much, as this movie is just so poor. The story might be the outest romantic little so	neg	10006_2.txt
8	test	"Tokyo Eyes" tells of a 17 year old Japanese girl who falls in love with a man being hunted by her big bro who is a cop. This lame flick is about 50% filler and 50% talk, talk, and more talk	neg	10007_4.txt
9	test	Wealthy horse ranchers in Buenos Aires have a long-standing no-holding policy with the Crawford of Manhattan, but what happens when the mustachioed Latin son falls for a certain Co	neg	10008_4.txt
10	test	Cage plays a drunk and gets high critical praise. Elizabeth Shue Actually has to do a love scene with the most unattractive and overrated piece of dung flesh in Hollywood. I literally vom	neg	10009_3.txt
11	test	First of all, I would like to say that I am a fan of all of the actors that appear in this film and at the time that I rented it, I wanted to like it. I think that the main reason that I was	neg	10010_3.txt
12	test	So tell me... what serious boozier drinks Budweiser? How many suicidally obsessed drinkers house a fully stocked and barely touched range of drinks in their lonely motel room that a mi	neg	10010_2.txt
13	test	A big disappointment for what was touted as an incredible film. Incredibly bad. Very pretentious. It would be nice if just once someone would create a high profile role for a young woman	neg	10011_1.txt
14	test	This film is absolutely appalling and awful. It's not low budget, it's a no budget film that makes Ed Wood's movies look like art. The acting is abysmal but sets and props are worse than a	neg	10012_1.txt
15	test	Here's a decidedly average Italian post-apocalyptic take on the hunting/killing humans for sport theme ala The Most Dangerous Game, Turkey Shoot, Gynkata and The Running Man. 	neg	10013_4.txt
16	test	At the bottom end of the apocalypse movie scale is this piece of pish called "The Final Executioner"... at least where I come from. A bloke is trained by an ex-cop to seek vengeance on th	neg	10014_2.txt
17	test	Earth has been destroyed in a nuclear holocaust. Well, parts of the Earth, because somewhere in Italy, a band of purebred survivors--those without radioactive contamination--are holed	neg	10015_4.txt
18	test	Many people are standing in front of the house n some women are crying... Men standing in close groups and speaking in hushed up tone... a couple of guys come in and they are discou	neg	10016_3.txt
19	test	New York family is the last in their neighborhood to get a television set, which nearly ruins David Niven's marriage to Milzi Caynor. Bedroom comedy that rarely ventures into the bedroom	neg	10017_1.txt
20	test	The best thing about "The Prey" is the tag line...."It's not human and it's got an axe!" The movie itself is a padded shrinko....endless insect and wildlife shots make the viewer wanna die	neg	10018_1.txt
21	test	This is truly, without exaggerating, one of the worst Slasher movies ever made. I know, it came out in the 80's following a tendency started by "Friday the 13th". "The Prey" copies the for	neg	10019_1.txt
22	test	I'm a huge fan of both Emily Watson (Breaking The Waves) and Tim Wilkinson (Normal) and was amused to see them upstaged by Rupert Everett (Dellamorte Dellamore) in this shocking	neg	1001_4.txt
23	test	Sure, most of the slasher films of the 1980's were not worth the celluloid they were filmed on, but this video nightmare may well be the dullest produced. 	neg	10020_1.txt
24	test	I think that would have been a more appropriate title for this film, since it is padded to hell and back with stock footage of various bugs and animals. I recently found The Prey in its origi	neg	10021_3.txt
25	test	1990 was certainly a year for bad backwoods slasher movies. "Friday The 13th" and "The Burning" may have been the best ones but there were like always a couple of stinkers not far b	neg	10022_4.txt
26	test	Everything everyone has said already pretty much rings true when it comes to "The Prey". Endless nature footage, bad acting. Aside from these elements, this is a watchable film for slas	neg	10023_4.txt
27	test	Uhhh ... so, did they even have writers for this? Maybe I'm picky, but I like a little dialog with my movies. And, as far as slasher films go, just a sliver of character development will suffice,	neg	10024_3.txt

加载数据集

```
from keras.datasets import imdb

print('Loading data...')
(x_train, y_train), (x_test, y_test) = imdb.load_data(num_words=5000)
print(len(x_train), 'train sequences')
print(len(x_test), 'test sequences')
```

其中:

- 评论已预处理为整数序列
- 每个整数代表词典中单词位置
- num_words: 仅保留最常出现的5000个词, 舍弃低频单词

输出

```
Using TensorFlow backend.
Loading data...
25000 train sequences
25000 test sequences
```

准备数据

```
from keras.preprocessing import sequence

print('Pad sequences (samples x time)')
x_train = sequence.pad_sequences(x_train, maxlen=400)
x_test = sequence.pad_sequences(x_test, maxlen=400)
print('x_train shape:', x_train.shape)
print('x_test shape:', x_test.shape)
```

- 不能将整数序列直接输入神经网络
- 将列表转化为张量
- 使用pad_sequences对序列进行填充
- 补0至长度相同，等于maxlen

输出

```
Pad sequences (samples x time)
x_train shape: (25000, 400)
x_test shape: (25000, 400)
```

建立模型

```
model = Sequential()
# Add an embedding layer
model.add(Embedding(input_dim=5000, output_dim=50, input_length=maxlen))
model.add(Dropout(0.2))
# Add a Convolution1D
model.add(Conv1D(filters=250, kernel_size=3, padding='valid', activation='relu', strides=1))
# Use max pooling:
model.add(GlobalMaxPooling1D())
# Add a full connected layer:
model.add(Dense(units=250))
model.add(Dropout(0.2))
model.add(Activation('relu'))
# Output layer with a sigmoid:
model.add(Dense(units=1))
model.add(Activation('sigmoid'))

model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])
model.fit(x_train, y_train, batch_size=32, epochs=5, validation_data=(x_test, y_test))
```


实验结果

```
Train on 25000 samples, validate on 25000 samples
Epoch 1/5
2019-10-16 16:34:46.034037: I tensorflow/core/platform/cpu_feature_guard.cc:142] Your CPU
supports instructions that this TensorFlow binary was not compiled to use: AVX2 FMA
25000/25000 [=====] - 42s 2ms/step - loss: 0.4074 - acc: 0.7994 -
val_loss: 0.2825 - val_acc: 0.8806
Epoch 2/5
25000/25000 [=====] - 40s 2ms/step - loss: 0.2317 - acc: 0.9063 -
val_loss: 0.2614 - val_acc: 0.8923
Epoch 3/5
25000/25000 [=====] - 39s 2ms/step - loss: 0.1632 - acc: 0.9386 -
val_loss: 0.2708 - val_acc: 0.8939
Epoch 4/5
25000/25000 [=====] - 40s 2ms/step - loss: 0.1074 - acc: 0.9609 -
val_loss: 0.3127 - val_acc: 0.8898
Epoch 5/5
25000/25000 [=====] - 38s 2ms/step - loss: 0.0768 - acc: 0.9721 -
val_loss: 0.3380 - val_acc: 0.8857
```