

黄小猿

码猿3年

暂无认证

183

原创

821

粉丝

540

获赞

111

评论

88万+

访问

7248

积分

1278

收藏

2万+

周排名

6018

总排名

等级

TA的主页

私信

关注

搜博文文章

最新文章

2019年，把时间分给最靠谱的人和事！
(三月复盘)

为什么穷人越穷，富人越富？
理一下不同ROE的计算方式
因为它，有了支票有了钱。

小猿理财：巴菲特推荐了ETF指数基金，没推荐增强型指数基金，原来道理在这！

分类专栏

python

12篇

tensorflow

1篇

机器学习与数据挖掘

19篇

小技能

13篇

个人简介

10篇

学习方法

4篇

归档

2019

4月
2篇

1月
1篇

2018

12月
2篇

11月
11篇

10月
2篇

9月
1篇

8月
2篇

7月
13篇

6月
21篇

5月
14篇

4月
4篇

3月
2篇

2月
1篇

热门文章

循环神经网络（RNN）原理通俗解释
140751

让你看懂聚类分类
118406

主题模型（LDA）（一）通俗理解与简单应用
51650

最优化理论与凸优化到底是干嘛的？
43678

python数据结构之链表（linked list）
40892

最新评论

详细解读Spatial Trans...
qq_23656689 回复 qq_23656689：好劲，和我现在的一样，谢谢！

详细解读Spatial Trans...
qq_23656689 回复 u014734471：这里应该有点问题x'这里是对theta的第一行的向...，

什么是思维模型？什么是原则？
AGS-wangdsh：是这样吧：原则：倾向于于个体的经验积累，但模型更倾向于这个...，

详细解读Spatial Trans...
yanglanshi09 回复 weixin_42808812：V是没有像素信息的，只有坐标位置信息，...，
数据挖掘师八大能力之一：收集信息的...
stapa 回复 vaae_musician：买了，可靠...

亿维云

高防服务器最高
防御1000G攻击

亿速云高防服务器，
20+行业领袖强烈推荐 免
备案服务器低延时CN2高
速带宽即买即用

目录

1.RNN怎么来的？

2.RNN的网络结构及原理

3.RNN的改进1：双向RNN

4.RNN的改进2：深层双向RNN

4.1 Pyramidal RNN

5.RNN的训练-BPTT

6.RNN与CNN的结合应用：看图说话

7.RNN项目练手

循环神经网络（RNN）原理通俗解释

原创

置顶

黄小猿

2017-11-30 15:15:37

140894

收藏 269

版权

分类专栏：

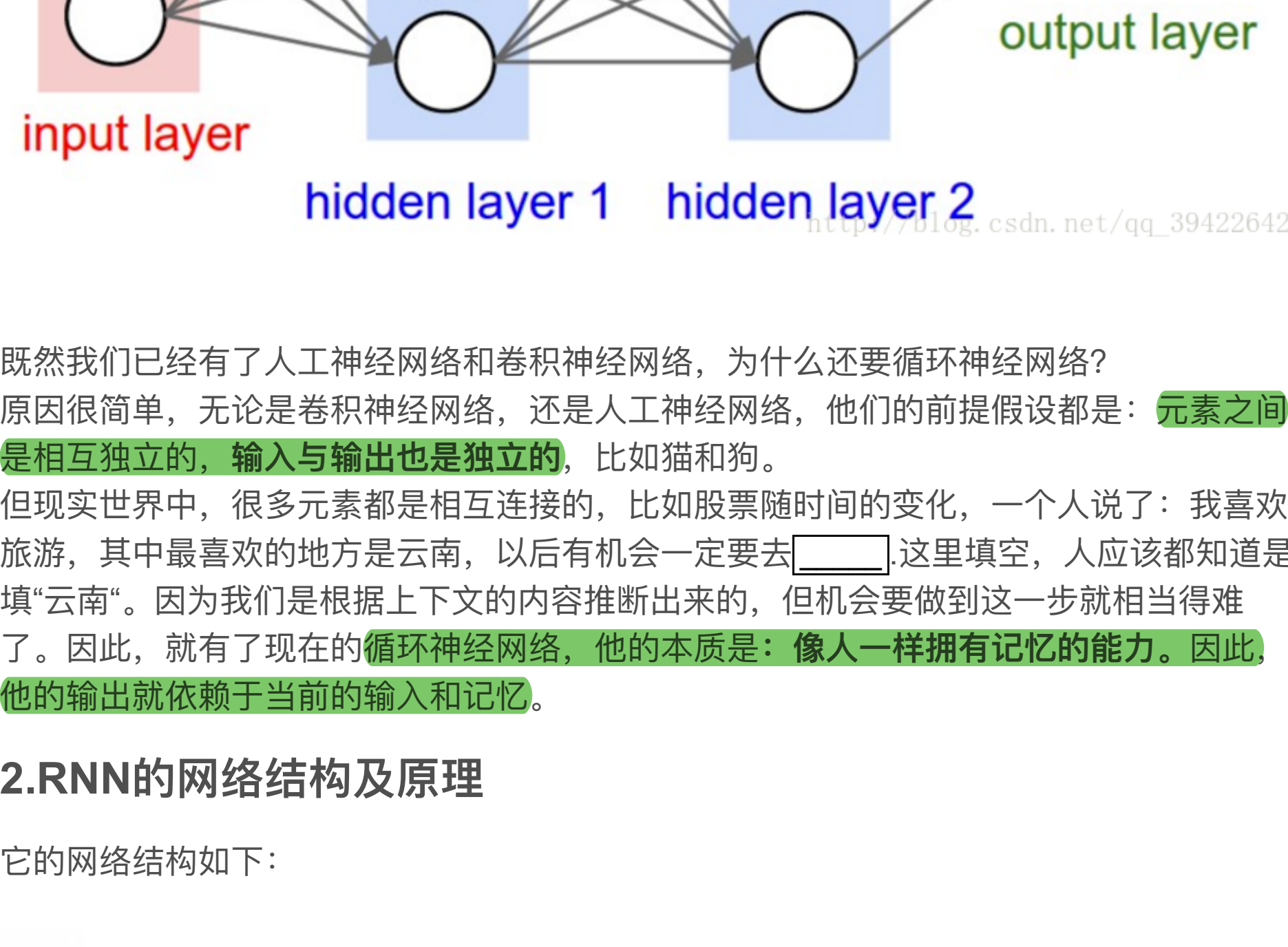
NLP

- 1.RNN怎么来的？
- 2.RNN的网络结构及原理
- 3.RNN的改进1：双向RNN
- 4.RNN的改进2：深层双向RNN
- 4.1 Pyramidal RNN
- 5.RNN的训练-BPTT
- 6.RNN与CNN的结合应用：看图说话
- 7.RNN项目练手

1.RNN怎么来的？

循环神经网络的应用场景比较多，比如暂时能写论文，写程序，写诗，但是，（总是会有但是的），但是他们现在还不能正常使用，学习出来的东西没有逻辑，所以要想真正让它更有用，路还很长。

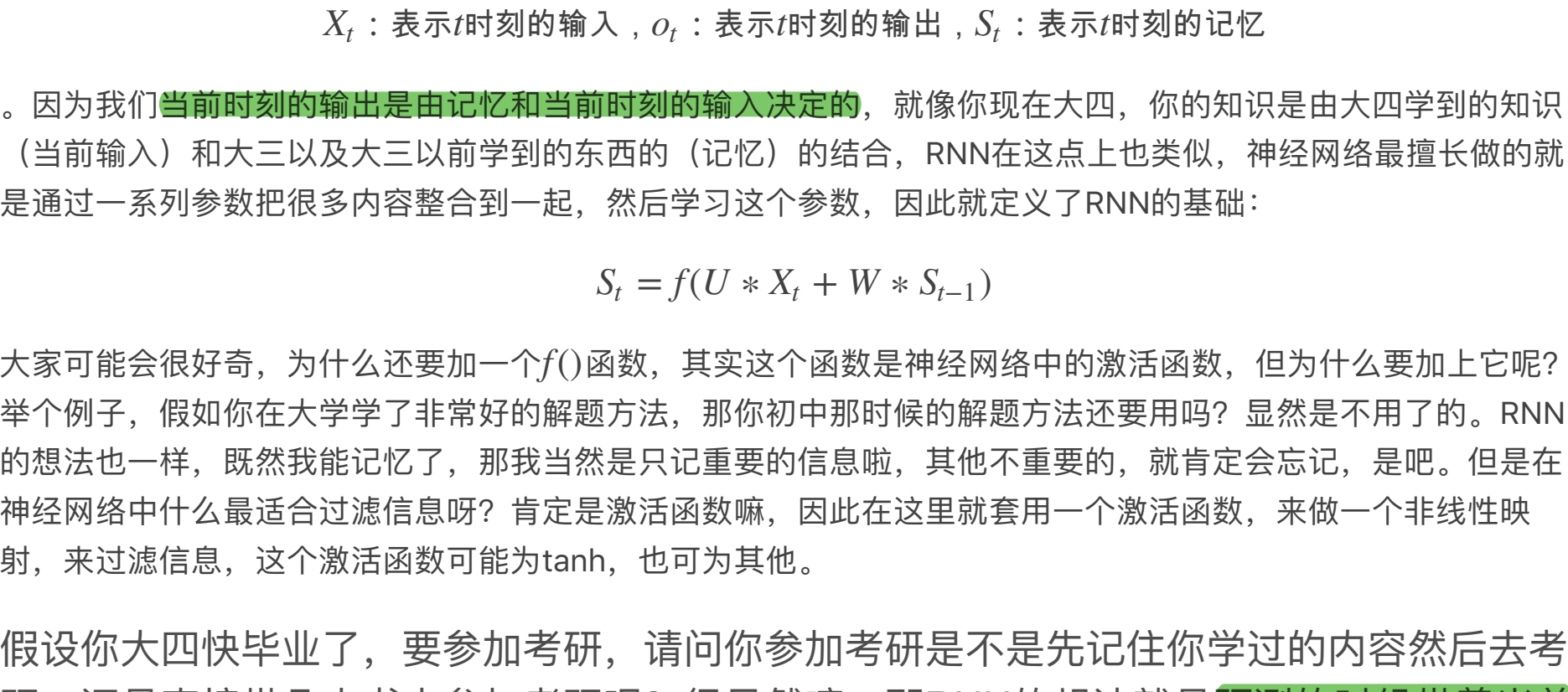
这是一般的神经网络应该有的结构：



既然我们已经有了人工神经网络和卷积神经网络，为什么还要循环神经网络？原因很简单，无论是卷积神经网络，还是人工神经网络，他们的前提假设都是：元素之间是相互独立的，输入与输出也是独立的。比如猫和狗。但现实世界中，很多元素都是相互连接的，比如股票随时间的变化，一个人说了：我喜欢旅游，其中最喜欢的地方是云南，以后有机会一定要去_____这里填空。人应该都知道是填“云南”。因为我们是在根据上下文的内容推断出来的，但机会要做到这一步就相当得难了。因此，就有了现在的循环神经网络，他的本质是：像人一样拥有记忆的能力。因此他的输出就依赖于当前的输入和记忆。

2.RNN的网络结构及原理

它的网络结构如下：



其中每个圆圈可以看作是一个单元，而且每个单元做的事情也是一样的，因此可以折叠呈左半图的样子。用一句话解释RNN，就是一个单元结构重复使用。

RNN是一个序列到序列的模型，假设 x_{t-1}, x_t, x_{t+1} 是一个输入：“我是中国”，那么 o_{t-1}, o_t 就应该对应是，“中国”这两个，预测下一个词最有可能是什么？就是 o_{t+1} 应该是人的概率比较大。

因此，我们可以做这样的定义：

$$X_t$$
：表示 t 时刻的输入， o_t ：表示 t 时刻的输出， S_t ：表示 t 时刻的记忆

。因为我们前时刻的输出是由记忆和当前时刻的输入决定的，就像你现在大四，你的知识是由大四学到的知识（当前输入）和高三以及大三之前学到的东西的（记忆）的结合，RNN在这点上类似，神经网络最擅长的就是通过一系列参数把很多内容整合到一起，然后学习这个参数，因此就定义了RNN的基础：

$$S_t = f(U * X_t + W * S_{t-1})$$

大家可能会很好奇，为什么还要加一个 $f()$ 函数，其实这个函数是神经网络中的激活函数，但为什么要加上它呢？举个例子，假如你在大学学了非常好的解题方法，那你初中那时候的解题方法还用吗？显然是不用了。RNN的想法也一样，既然我能记住了，那我当然是只记重要的信息啦，其他不重要的，就肯定会忘记，是吧，但是在神经网络中什么量适合过速信息呀？肯定是激活函数嘛，因此在这里就套用一个激活函数，来做一个非线性映射，来过滤信息，这个激活函数可能为tanh，也可为其他。

假设你大四快毕业了，要参加考研，请问你参加考研是不是先记住你学过的内容然后去考研，还是直接带几本书去参加考研呢？很显然嘛，那RNN的想法就是：预测的时候带着当前时刻的记忆 S_t 去预测。假如你要预测“我是中国”的下一个词出现的概率，这里已经很显然了，运用softmax来预测每个词出现的概率再合适不过了，但预测不能直接用一个矩阵来预测呀，所有预测的时候还要带一个权重矩阵 V 用公式表示为：

$$o_t = softmax(V * S_t)$$

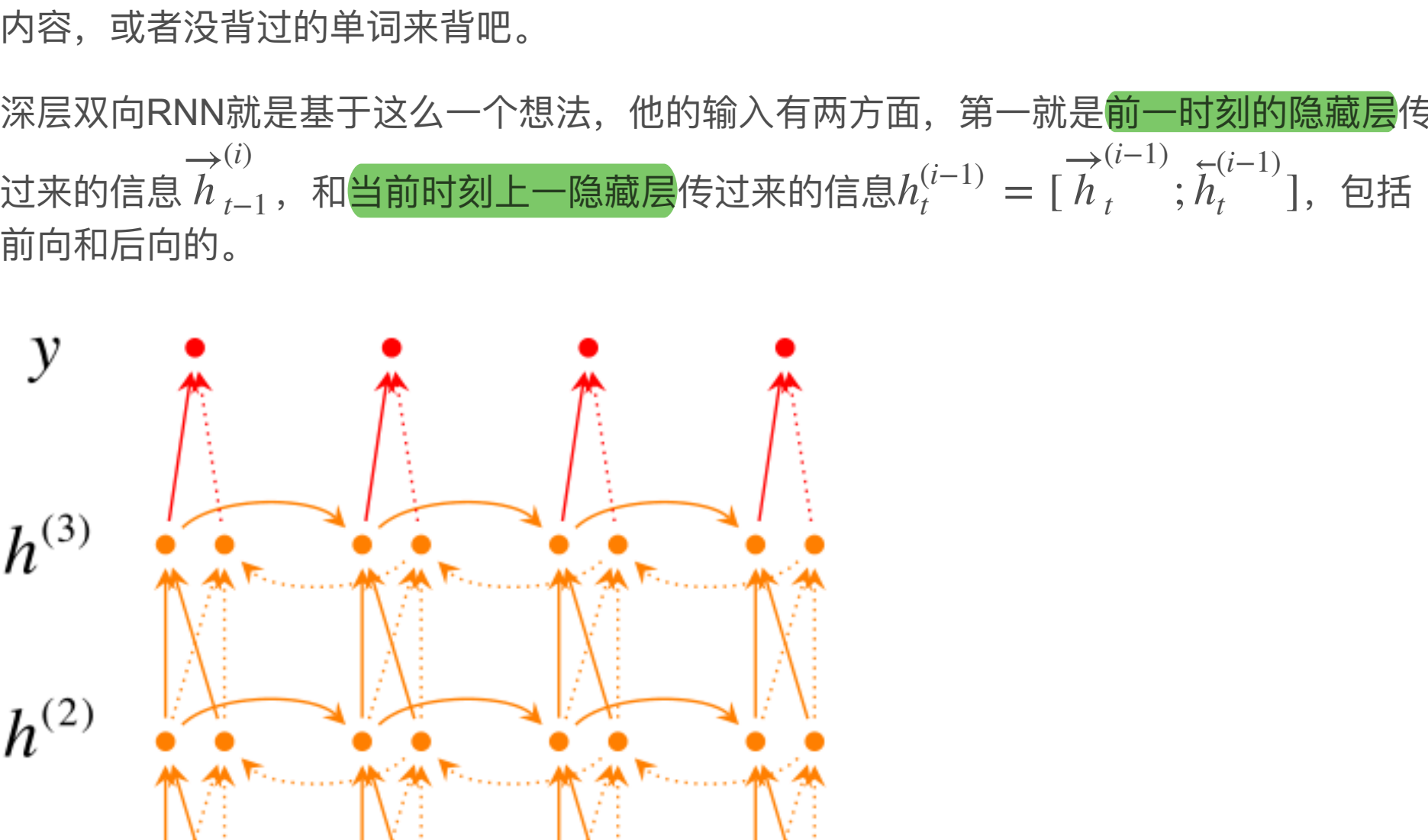
其中 o_t 就表示时刻的输出。

RNN中的结构细节：

- 1.可以把 S_t 当作隐状态，捕捉了之前时间点上的信息。就像你去考研一样，考的时候记住了你能记住的所有信息。
2. o_t 是由当前时间以及之前所有的记忆得到的。就是你考研之后做的考试卷子，是用你的记忆得到的。
- 3.很可惜的是， S_t 并不能捕捉之前所有时间点的信息。就像你考研不能记住所有的英语单词一样。
- 4.和卷积神经网络一样，这里的网络中每个cell都共享了一组参数（ U, V, W ），这样就能极大的降低计算量了。
5. o_t 在很多情况下都是不存在的，因为很多任务，比如文本情感分析，都是只关注最后的结果的。就像考研之后选择学校，学校不会管你到底怎么努力，怎么心酸的准备考研，而只关注你最后考了多少分。

3.RNN的改进1：双向RNN

在有些情况，比如有一部电视剧，在第三集的时候才出现的人物，现在来预测一下在第三集中出现的人物名字，你用前面两集的内容是预测不出来的，所以需要用到第四，第五集的内容来预测第三集的内容，这就是双向RNN的想法。如图是双向RNN的图解：



从前往后： $\vec{S}_t^f = f(U^f * X_t + W^f * S_{t-1}^f + b^f)$

从后往前： $\vec{S}_t^b = f(U^b * X_t + W^b * S_{t-1}^b + b^b)$

输出： $o_t = softmax(V * [\vec{S}_t^f; \vec{S}_t^b])$

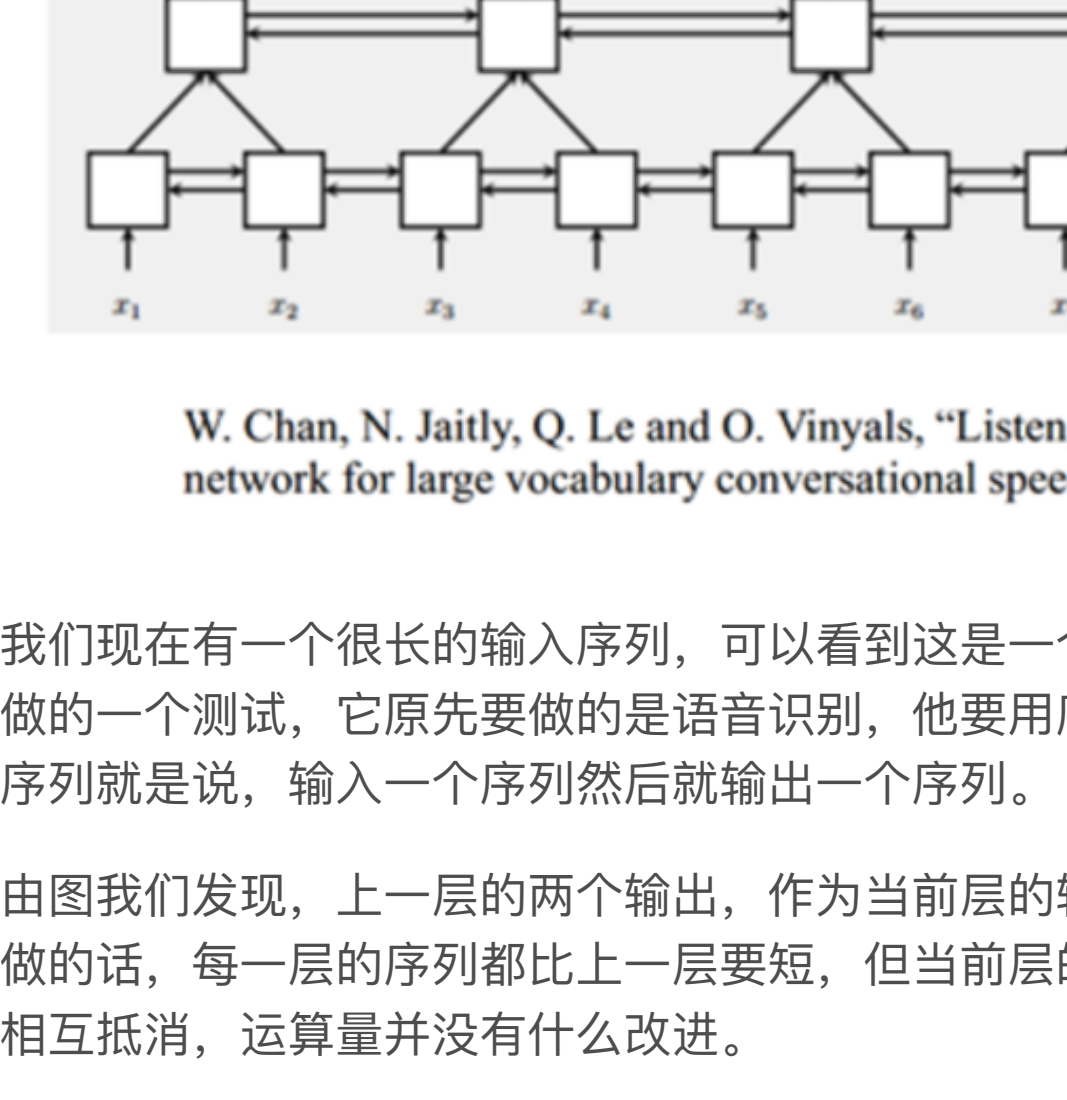
这里的 $[\vec{S}_t^f; \vec{S}_t^b]$ 做的就是一个拼接，如果他们都是1000x1维的，拼接在一起就是1000x2维的了。

双向RNN需要的内存是单向RNN的两倍，因为在同一时间点，双向RNN需要保存两个方向的权重参数，在分类的时候，需要同时输入两个隐藏层输出的信息。

4.RNN的改进2：深层双向RNN

深层双向RNN与双向RNN相比，多了几个隐藏层，因为他的想法是很多信息记一次记不下来，比如你去考研，复习考研英语的时候，背英语单词一定不会就着一次就记住了所有要考的考研单词吧，你应该也是带着先前几次背过的单词，然后选择那些背过，但不熟的内容，或者没背过的单词来背吧。

深层双向RNN就是基于这么一个想法，他的输入有两方面，第一就是前一刻的隐藏层传过来的信息 $\vec{h}_{t-1}^{(i)}$ ，和当前时刻上一隐藏层传过来的信息 $\vec{h}_t^{(i-1)} = [\vec{h}_t^{(i-1)}; \vec{h}_{t-1}^{(i-1)}]$ ，包括前向和后向的。



我们用公式来表示是这样的：

$$\vec{h}_t^{(i)} = f(\vec{W}^{(i)} h_t^{(i-1)} + \vec{V}^{(i)} \vec{h}_{t-1}^{(i)} + \vec{b}^{(i)})$$

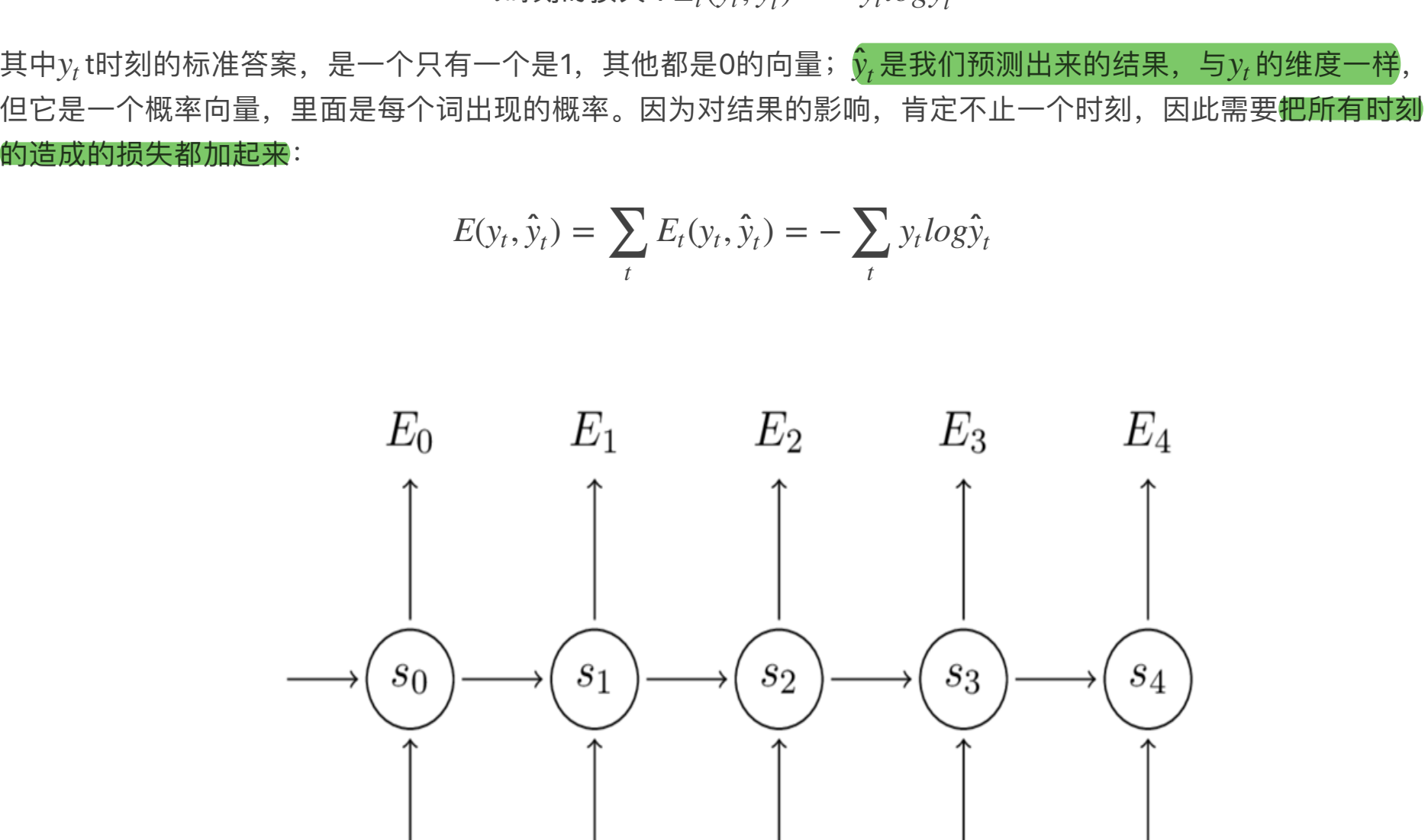
$$\vec{h}_t^{(i)} = f(\vec{W}^{(i)} h_t^{(i-1)} + \vec{V}^{(i)} \vec{h}_{t+1}^{(i)} + \vec{b}^{(i)})$$

然后再利用最后一层来进行分类，分类公式如下：

$$\hat{y}_t = g(Uh_t + c) = g(U[\vec{h}_t^{(L)}; \vec{h}_t^{(L)}] + c)$$

4.1 Pyramidal RNN

其他类似的网络还有Pyramidal RNN：



W. Chan, N. Jaitly, Q. Le and O. Vinyals, "Listen, attend and spell: A neural network for large vocabulary conversational speech recognition," ICASSP, 2016

我们现在有一个很长的输入序列，可以看到这是一个双向的RNN，上图是谷歌的W.Chan做的一个测试，它原先要做的是语音识别，他要用序列到序列的模型做语音识别，序列到序列就是说，输入一个序列然后输出一个序列。

由图我们发现，上一层的两个输出，作为当前层的输入，如果是非常长的序列的话，这样做的话，每一层的序列都比上一层要短，但当前层的输入 $f(x)$ 也会随之增多，貌似看起来相互抵消，运算量并没有什么改进。

但我们知道，对于一层来说，它是从前往后转的，比如要预测一个股市的变化，以天为单位，假如要预测明天的股市变化，你就用今天，以及今天之前的所有数据，我们暂时无法只用昨天的数据，不用今天的数据，预测明天的数据，也即是说，预测必须具有连续性。

但每一层的 f 运算是可以并行的，从这个角度来看，运算量还是可以接受的，特别是在原始输入序列较短的时候还是有优势的。

5.RNN的训练-BPTT

如前面我们讲的，如果要预测时刻的输出，我们必须要利用上一时刻（t-1）的记忆和当前时刻的输入，得到时刻的输出：

$$s_t = \tanh(Ux_t + WS_{t-1})$$

然后利用当前时刻的记忆，通过softmax函数输出每个词出现的概率：

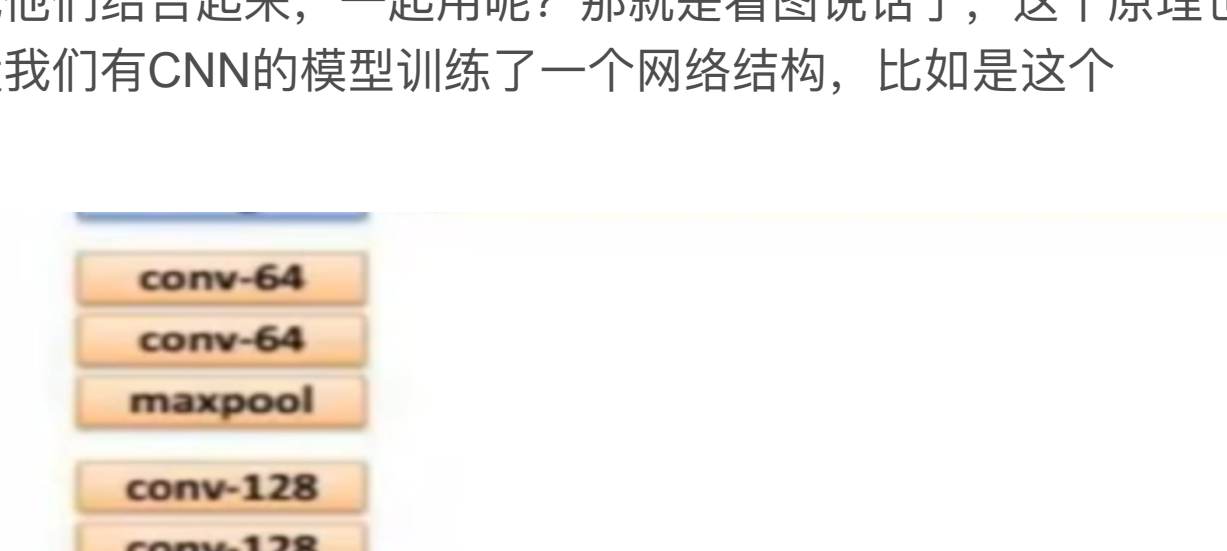
$$\hat{y}_t = softmax(Vs_t)$$

为了找出模型是最好的参数，U，W，V，我们就要知道当前参数得到的结果怎么样，因此就要定义我们的损失函数，用交叉熵损失函数：

$$t$$
时刻的损失： $E_t(y_t, \hat{y}_t) = -y_t \log \hat{y}_t$

其中 y_t 时刻的标准答案，是一个只有一个量1，其他都是0的向量： y_t 是我们预测出来的结果，与 y_t 的维度一样，但它是一个概率向量，里面是每个词出现的概率，因为对结果的影响，肯定不止一个时刻，因此需要把所有时刻造成的损失都加起来：

$$E(y_t, \hat{y}_t) = \sum_t E_t(y_t, \hat{y}_t) = - \sum_t y_t \log \hat{y}_t$$



如图所示，你会发现每个cell都会有一个损失，我们已经定义好了损失函数，接下来就是熟悉的一步了，那就是根据损失函数使用SGD来求解最优参数，在CNN中使用反向传播BP算法来求解最优参数，但在RNN就要用到BPTT，它和BP算法的本质区别，也是CNN和RNN的本质区别：CNN没有记忆功能，它的输出仅依赖于输入，但RNN有记忆功能，它的输出不仅依赖于当前输入，还依赖于当前的记忆，这个记忆是序列到序列的，也就是当前时刻收到上一时刻的影响，比如股市的变化。

因此，在对参数求导的时候，对当前时刻求导，一定会涉及前一刻，我们用例子看一下：



假设我们对 E_3 的 W 求偏导：它的损失首先来源于预测的输出 \hat{y}_3 ，预测的输出又是来源于当前时刻的记忆 s_3 ，当前的记忆又是来源于当前的输出和截止到上一时刻的记忆：

$$s_3 = \tanh(Ux_3 + WS_2)$$

因此，根据链式法则可以有：

$$\frac{\partial E_3}{\partial W} = \frac{\partial E_3}{\partial \hat{y}_3} \frac{\partial \hat{y}_3}{\partial s_3} \frac{\partial s_3}{\partial W}$$

但是，你会发现， $s_2 = \tanh(Ux_2 + WS_1)$ ，也就是 s_2 里面的函数还包含了 W ，因此，这个链式法则还没到头，就像图上面的那样，所以真正的链式法则应该是这样的：

$$\frac{\partial E_3}{\partial W} = \sum_{k=0}^3 \frac{\partial E_3}{\partial \hat{y}_3} \frac{\partial \hat{y}_3}{\partial s_3} \frac{\partial s_3}{\partial s_k} \frac{\partial s_k}{\partial W}$$

我们要是把当前时刻造成的损失，和以往每个时刻造成的损失加起来，因为我们每个时刻都带有了权重参数 W ，和以往的损失不同，一般的神经网络，比如人工神经网络，参数是不同共享的，但在循环神经网络，和CNN一样，设立了参数共享机制，来降低模型的计算量。

6.RNN与CNN的结合应用：看图说话

在图像处理中，目前做的最好的是一起用呢，而自然语言处理中，表现比较好的是RNN，因此我们能否把他们结合起来，成为CNN，那就是看图说话了，这个原理也比较简单，举个小栗子：假设我们有CNN的模型训练了一个网络结构，比如是这个



最后我们不是要分类嘛，那在分类前，是不是已经拿到了图像的特征呀，那我们能不能把图像的特征拿出来，放到RNN的输入里，让他去学习呢？

之前的RNN是这样的：

$$S_t = \tanh(U * X_t + W * S_{t-1})$$

我们把图像的特征加在面前，可以得到：

$$S_t = \tanh(U * X_t + W * S_{t-1} + V * X)$$

其中的 X 就是图像的特征。如果用的是上面的CNN网络， X 应该是一个4096x1的向量。

注：这个公式只在第一步做，后面每次更新就没有 V 了，因为给RNN数据只在第一次迭代的时候做。

7.RNN项目练手

RNN可以写歌词，写诗等，这有个项目可以玩玩，还不错。
Tensorflow实现RNN

大白话讲循环神经网络RNN-从此登上所有知... huacha_的博文 @ 17+
本系列讲解循环神经网络RNN和LSTM的所有知识点。学员本课程将对于RNN和LSTM的理论知识有清晰的认识... 8-10
深度学习笔记——RNN（LSTM、GRU、双向RNN）学习总结 mpk_no1的博文 @ 5万+
本文是关于RNN和RNN的变种LSTM、GRU以及BiRNN的学习总结。

深度学习中-RNN循环神经网络-理论+图解+Python代码部分) huacha_的博文 @ 17+
目录前言一、普通神经网络二、循环神经网络1.前向传播2.反向传播算法BPTT3.自然语言处理-RNN 4.难点问题... 8-15
如何从RNN起步，一步步通俗理解LSTM 结构之算法之道 @ 17+
如何从RNN起步，一步步通俗理解LSTM前言提到LSTM，之前学过的同学可能最先想到的是Christopher Olah的...
RNN了了不起的战队-CSDN博客_rnn 8-20
RNN(Recurrent Neural Network)是一类用于处理序列数据的神经网络。首先我们要明确什么是序列数据,抽取百度...
RNN-硬着头皮编程的王大爷-CSDN博客_rnn 8-15
引入RNN(Recurrent Neural Network)就很好地解决了输入之间相互依赖的问题。2. RNN 2.1 RNN结构体 RNN...
深度学习之RNN(循环神经网络) 笨拙的右脑的博文 @ 9090
一 RNN概述 前面我们叙述了BP算法, CNN算法, 那么为什么还会有RNN呢? 什么是RNN, 它到底有什么不同之...
循环神经网络(RNN, Recurrent Neural Networks)介绍 我和刘逍遥的梦--- @ 43万+
循环神经网络(RNN, Recurrent Neural Networks)介绍 这篇文章很多内容是参考: http://www.wildml.com/201...
RNN介绍-较易懂_xiaocao9903的专栏-CSDN博客_rnn 8-17
2017 12月 5篇 11月 14篇 7月 1篇 5月 2篇 4月 1篇 3月 6篇 2月 2篇 2016 12月 1篇 11月 19篇 10月 11篇 9月 9篇...
RNN详解(Recurrent Neural Network, bestrivern)的差-CSDN博客_rnn 8-16
一,概述RNN用于处理序列数据。在传统的神经网络模型中,输入到输出是逐层串行,层与层之间是全连接...
循环神经网络 (RNN) 入门 二师兄 @ 9090
之前学习了卷积神经网络 (CNN) , 在这里再简单介绍一下卷积神经网络 的原理。一个典型的卷积神经网络为: ...
循环神经网络 (RNN) 原理以及实战 8-11
一 RNN的网络结构及原理 2. RNN的改进 2.1 双向RNN 2.2 深层双向RNN 五 Pyramidal RNN 六 RNN的训练-BPTT 七...
RNN 人工智能_yumi_huang的博文-CSDN博客 8-21
“ 人工智... https://blog.csdn.net/haoyi1995/article/details/80572098笔记: 基础的神经网络只在层与层之间建立了...
RNN_诗露的专栏-CSDN博客_rnn_cnds 8-15
文章目录循环神经网络与前馈神经网络的异同点RNN的展开模型基于时间的反向传播算法调整sw_3所需的模型...
深度学习基础 (人工神经网络、CNN、RNN、Istm) 02-09
人工神经网络、CNN、RNN、Istm
循环神经网络 天空飘来五个字 @ 8873
一、循环神经网络简介循环神经网络的来源是为了刻画一个序列当前的输出与之前信息的关系,从网络结构上, ...
RNN_欢迎来到小丁的技术空间-CSDN博客 3-27
这个链接对于公式讲解的非常清楚,以及RNN的各种变形
RNN介绍-阿华Go从现在开始的博文-CSDN博客 6-23
RNN可以被看做是同一神经网络的多次赋值,每个神经网络模块会消息传递给下一个,所以,如果我们得这个微...
©2020 CSDN 皮肤主题: 书香水墨 设计师: CSDN官方博客 返回首页