

面试常问的深度学习(DNN、CNN、RNN)的相关问题

2017年07月24日 16:46:40

阅读量：3790

神经网络

的学习就是学习如何利用矩阵的线性变换加激活函数的非线性变换，将原始输入空间投向线性可分/稀疏的空间去分类/回归。**增加节点数：增加维度，即增加线性转换能力。增加层数：增加激活函数的次数，即增加非线性转换次数。**

对卡在局部极小值的处理方法：

1.调节步伐：调节学习速率，使每一次的更新“步伐”不同；**2.优化起点：**合理初始化权重（weights initialization）、预训练网络（pre-train），使网络获得一个较好的“起始点”，如最右侧的起始点就比最左侧的起始点要好。常用方法有：高斯分布初始权重（Gaussian distribution）、均匀分布初始权重（Uniform distribution）、Glorot 初始权重、He初始权、稀疏矩阵初始权重（sparse matrix）。

浅层VS深层：

浅层神经网络可以模拟任何函数，但数据量的代价是无法接受的。深层解决了这个问题。相比浅层神经网络，深层神经网络可以用**更少的数据量**来学到更好的拟合。深层的前提是：空间中的元素可以由迭代发展而来的。

防止过拟合：

L2正则化，**Dropout**(若规律不是在所有样本中都存在,则dropout会删除这样的规律)，每个epoch之后shuffle训练数据，设置**early-stopping**。加**Batch Normalization**(BN首先是把所有的samples的统计分布标准化，降低了batch内不同样本的差异性，然后又允许batch内的各个samples有各自的统计分布)，BN最大的优点为允许网络使用较大的学习速率进行训练，加快网络的训练速度（减少epoch次数），提升效果。

为何使用Batch Normalization：

若用**多个梯度的均值**来更新权重的批量梯度下降法可以用相对少的训练次数遍历完整个训练集，其次可以使更新的方向更加贴合整个训练集，避免单个噪音样本使网络更新到错误方向。然而也正是因为平均了多个样本的梯度，许多样本对神经网络的贡献就被其他样本平均掉了，相当于在每个epoch中，训练集的样本数被缩小了。batch中每个样本的差异性越大，这种弊端就越严重。一般的解决方法就是在每次训练完一个epoch后，将训练集中样本的顺序打乱再训练另一个epoch，不断反复。这样重新组成的batch中的样本梯度的平均值就会与上一个epoch的不同。而这显然增加了训练的时间。同时因为没办法保证每次更新的方向都贴合整个训练集的大方向，只能使用较小的学习速率。这意味着训练过程中，一部分steps对网络最终的更新起到了促进，一部分steps对网络最终的更新造成了干扰，这样“磕磕碰碰”无数个epoch后才能达到较为满意的结果。

为了解决这种“不效率”的训练，BN首先是把所有的samples的统计分布标准化，降低了batch内不同样本的差异性，然后又允许batch内的各个samples有各自的统计分布。

- 为什么神经网络高效：**并行的先验知识使得模型可用线性级数量的样本学习指数级数量的变体
- 学习的本质是什么：**将变体拆分成因素和知识（Disentangle Factors of Variation）
 - 为什么深层神经网络比浅层神经网络更高效：**迭代组成的先验知识使得样本可用于帮助训练其他共用同样底层结构的样本。
 - 神经网络在什么问题上的不具备优势：**不满足并行与迭代先验的任务
- 非迭代：**该层状态不是由上层状态构成的任务（如：很深的CNN因为有max pooling，信息会逐渐丢失。而residual network再次使得迭代的先验满足）

CNN:

1) 卷积：对图像元素的矩阵变换，是提取图像特征的方法，多种卷积核可以提取多种特征。一个卷积核覆盖的原始图像的范围叫做感受野（权值共享）。一次卷积运算(哪怕是多个卷积核)提取的特征往往是局部的，难以提取出比较

2) 池化：降维的方法，按照卷积计算得出的特征向量维度大的惊人，不但会带来非常大的计算量，而且容易出现过拟合，解决过拟合的办法就是让模型尽量“泛化”，也就是再“模糊”一点，那么一种方法就是把图像中局部区域的特征做一个平滑压缩处理，这源于局部图像一些特征的相似性(即局部相关性原理)。

3) 全连接：softmax分类

训练过程：

卷积核中的因子($\times 1$ 或 $\times 0$)其实就是需要学习的参数，也就是卷积核矩阵元素的值就是参数值。一个特征如果有9个值，1000个特征就有900个值，再加上多个层，需要学习的参数还是比较多的。

CNN的三个优点：

sparse interaction(稀疏的交互)，parameter sharing(参数共享)，equivalent representation(等价表示)。适合于自动问答系统中的答案选择模型的训练。

CNN与DNN的区别：

DNN的输入是向量形式，并未考虑到平面的结构信息，在图像和NLP领域这一结构信息尤为重要，例如识别图像中的数字，同一数字与所在位置无关(换句话说任一位置的权重都应相同)，CNN的输入可以是tensor，例如二维矩阵，通过filter获得局部特征，较好的保留了平面结构信息。

filter尺寸计算：Feature Map的尺寸等于 $(input_size + 2 * padding_size - filter_size) / stride + 1$

RNN:

1. 为什么具有记忆功能？

这个是在RNN就解决的问题，就是因为有递归效应，上一时刻隐层的状态参与到了这个时刻的计算过程中，直白一点呢的表述也就是选择和决策参考了上一次的状态。

2. 为什么LSTM记的时间长？

因为特意设计的结构中具有CEC的特点，误差向上一个状态传递时几乎没有衰减，所以权值调整的时候，对于很长时间之前的状态带来的影响和结尾状态带来的影响可以同时发挥作用，最后训练出来的模型就具有较长时间范围内的记忆功能。

误差回传的主力还是通过了Memory Cell而保持了下来。所以我们现在用的LSTM模型，依然有比较好的效果。

最后整个梳理一下误差回传的过程，误差通过输出层，分类器，隐层等进入某个时刻的Block之后，先将误差传递给了Output Gate和Memory Cell两个地方。

到达输出门的误差，用来更新了输出门的参数w，到达Memory Cell之后，误差经过两个路径，

1是通过这个cell向前一个时刻传递或者更前的时刻传递，

2是用来传递到input gate和block的输入，用来更新了相应的权值(注意！不会经过这里向前一个时刻传递误差)。

最关键的问题就是，这个回传的算法，只通过中间的Memory Cell向更前的时刻传递误差。

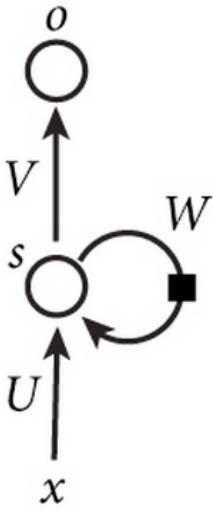
- 传统的RNN总是用“覆写”的方式计算状态： $S_t = f(S_{t-1}, x_t)$ ，其中 $f(\cdot)$ 表示仿射变换外面在套一个Sigmoid， x_t 表示输入序列在时刻 t 的值。根据求导的链式法则，这种形式直接导致梯度被表示为连乘积的形式，以致于造成梯度消失——粗略的说，很多个小于1的项连乘就很快逼近零。

- 现代的RNN(包括但不限于使用LSTM单元的RNN)使用“累加”的形式计算状态：

$$S_t = \sum_{\tau=1}^t \Delta S_{\tau}, \text{ 其中的 } \Delta S_{\tau} \text{ 显示依赖序列输入 } x_{\tau}. \text{ 稍加推导即可发现，这种累加形式导致导}$$

数也是累加形式，因此避免了梯度消失。

在RNN中U、V、W的参数都是共享的，也就是只需要关注每一步都在做相同的事情，只是输入不同，这样来降低参数个数和计算量。



RNN特点：

时序长短可变（只要知道上一时刻的隐藏状态 h_{t-1} 与当前时刻的输入 x_t ，就可以计算当前时刻的隐藏状态 h_t 。并且由于计算所用到的 W_{xh} 与 W_{hh} 在任意时刻都是共享的。递归网络可以处理任意长度的时间序列）顾及时间依赖，未来信息依赖（双向递归）

RNN主要包括LSTM,GRU

GRU对LSTM做了两个大改动：

1. 将输入门、遗忘门、输出门变为两个门：更新门（Update Gate）和重置门（Reset Gate）。
2. 将单元状态与输出合并为一个状态：。

GRU只用了两个gates，将LSTM中的输入门和遗忘门合并成了更新门。并且并不把线性自更新建立在额外的memory cell上，而是直接线性累积建立在隐藏状态上，并靠gates来调控。

版权声明：本文为博主原创文章，未经博主允许不得转载。 <https://blog.csdn.net/xwd18280820053/article/details/76026523>

文章标签：深度学习 cnn RNN

个人分类：deep learning 神经网络

想对作者说点什么？

CNN(卷积神经网络)、RNN(循环神经网络)、DNN(深度神经网络)概念区分理解

1、**相关知识**从广义上来说，NN（或是更美的**DNN**）确实可以认为包含了**CNN**、**RNN**这些具体的变种形式。有很多人认为，它们并没有可比性，或是根本没必要放在一起比较。在实际应用中，所谓的深度神经网络**DNN**...

Eddy_zheng 2016-02-29 10:21:34 阅读数：55578

CNN(卷积神经网络)、RNN(循环神经网络)、DNN(深度神经网络)的内部网络结构有什么区别？

CNN(卷积神经网络)、**RNN**(循环神经网络)、**DNN**(深度神经网络)的内部网络结构有什么区别？从广义上来说，NN(或是更美的**DNN**)确实可以认为包含了**CNN**、**RNN**这些具体的变种形式。在实际应用中...

Real_Myth 2016-12-21 09:13:03 阅读数：5325

2018-4-20

深度学习岗位面试问题整理笔记 - CSDN博客

本笔记主要问题来自以下两个问题,后续会加上我自己面试过程中遇到的问题。 深度...1: LSTM结构推导,为什么比RNN好? 答案:推导forget gate,input gate,cell state...

2018-5-11

农村有一宝,可治痔疮,可惜很少人知道!!

嘉泰·顶新

CNN、RNN、DNN区别

神经网络的来源 神经网络技术起源于上世纪五、六十年代,当时叫感知机(perceptron),包含有输入层、输出层和一个隐藏层。输入的特征向量通过隐藏层变换到达输出层,由输出层得到分类结果。...

lff1208 2017-08-30 15:33:13 阅读数:3302

面试经验(互联网,研究所,国企) - CSDN博客

这篇文章要介绍我在各个领域各种方式的面试经验,包括互联网、研究所、外企、国企...RNN与LSTM的matlab实现 1篇 分类效果的各种衡量标准 0篇 梯度下降 1篇 sv...

2018-2-24

面试常问的深度学习(DNN、CNN、RNN)的相关问题 - CSDN博客

举报内容: 面试常问的深度学习(DNN、CNN、RNN)的相关问题 举报原因: 色情 政治 抄袭 广告 招聘 骂人 其他 原文地址: 原因补充: 最多只允许输入30个字加...

2018-4-30

深度学习入门:一句话告诉你什么是神经网络(CNN,RNN,DNN)

神经网络技术起源于上世纪五、六十年代,当时叫感知机(perceptron),拥有输入层、输出层和一个隐含层。输入的特征向量通过隐含层变换达到输出层,在输出层得到分类结果。早期感知机的推动者是Rosen...

qq_39521554 2018-02-19 20:16:14 阅读数:156

DNN、CNN、RNN简析

首先,DNN、CNN、RNN可以一起比较。从广义上来说,NN(或是更美的DNN)确实可以认为包含了CNN、RNN这些具体的变种形式。在实际应用中,所谓的深度神经网络DNN,往往融合了多种已知的结构,包...

wuxiaoming1733 2016-05-13 17:26:54 阅读数:1418

BAT机器学习面试1000题系列(第1~325题) - CSDN博客

"结构之法",进入本博客),博客专注面试、算法、...LSTM结构推导,为什么比RNN好?深度学习 DL模型 难推导...

2018-5-13

大牛的深度学习面试经验 - CSDN博客

文章标签: 深度学习 面试经验 个人分类: 算法深度学习 想对作者说点什么? ...面试常问的深度学习(DNN、CNN、RNN)的相关问题 CNN: 1)卷积:对图像元素的矩阵...

2018-5-6

Deep Learning(深度学习)整理,RNN,CNN,BP

申明:本文非笔者原创,原文转载自:http://www.sigvc.org/bbs/thread-2187-1-3.html 4.2、初级(浅层)特征...

justdoithai 2016-04-16 21:47:16 阅读数:10439

如艺 · 顶新

深度学习常见面试题(更新中) - CSDN博客

笔面试题(1) 作者同类文章 X 1: LSTM结构推导,为什么比RNN好? 答案:推导forget gate,input gate,cell state, hidden information等的变化;因为LSTM有进有出...

2018-4-19

深度学习面试 - CSDN博客

(阿里面试)解释alpha狗: 解释resnet、优缺点以及适用范围: 解释inception net、优缺点以及适用范围: 解释RNN: GAN的公式以及发展历程: 会写公式知道变体 优...

2018-4-20

深度学习之 (DNN) 深度神经网络

(DNN) 深度神经网络 简介 DNN是指深度神经网络。与RNN循环神经网络、CNN卷积神经网络的别就是DNN特指全连接的神经元结构,并不包含卷积单元或是时间上的关联。 神经网络简史...

 lihaitao000 2016-05-03 21:42:32 阅读数 : 17561

答 : 那些深度学习《面试》你可能需要知道的

深度学习《面试》那些可能需要知道的内容答 : 那些深度学习《面试》你可能需要知道的 (知乎专栏)...

 nandehutugood_ 2017-11-26 11:55:31 阅读数 : 182

深度学习面试问题总结 - CSDN博客

RNN引入了循环的概念,但是在实际过程中却出现了初始信息随时间消失的问题,即长期...因为2017年就要回学校开始找工作了,必然免不了遇到各种面试问题,我先开个专栏记录...

2018-5-3

undefined

RNN介绍, 较易懂

原文 <http://www.jianshu.com/p/9dc9f41f0b29> Recurrent Neural Networks 人类并不是每时每刻都从一片空白的大脑开始他们的思...

 prom1201 2016-08-16 15:42:03 阅读数 : 59916

通俗理解RNN

全连接神经网络和卷积神经网络他们都只能单独的取处理一个个的输入,前一个输入和后一个输入是完全没有关系的。但是,某些任务需要能够更好的处理序列的信息,即前面的输入和后面的输入是有关系的。比如,当我们在理...

 qq_23225317 2017-09-04 11:14:42 阅读数 : 2190

深度学习面试题

为了准备面试,所以在网上搜集一些深度学习面试题,还有自己面试过程中遇到的一些问题吧。我自己面试的: 1 SVM推导, SVM多分类方法(1对1, 1对多, 多对多), lr loss function 推导...

 gdmzmj 2017-08-13 20:48:01 阅读数 : 8076

CNN笔记: 通俗理解卷积神经网络

通俗理解卷积神经网络 (cs231n与5月dl课程笔记) 1 前言 2012年我在北京组织过8期machine learning读书会,那时“机器学习”非常火,很多人都对其抱有巨大的热情。当我20...

 v_JULY_v 2016-07-02 22:14:50 阅读数 : 139672

CNN与常用框架

七月在线 5月深度学习班课程笔记——No.4 这一节内容比较丰富。主要介绍神经网络的每一个层级结构,理清每一层可能影响结果的参数。再介绍CNN的常用框架。 1 神经网络 1 神经网络——是

对于程序员来说，英语到底多重要

老司机教你一个数学公式秒懂天下英语



深度学习岗位面试问题整理笔记

转自<https://zhuanlan.zhihu.com/p/25005808> 更新几个面试被问到或者联想出来的问题，后面有时间回答 SGD 中 S(stochastic)代表什么 ...

 lj6052317 2017-10-15 17:15:29 阅读数：1354

深度学习面试题2018

1、问题：如何优化模型：加速收敛，避免overfit, 提升精度 ..? 答案：可以从以下几个参数开始：- batch size effect ; - learning rate effect...

 qq_27923041 2017-10-31 21:32:56 阅读数：2149

机器学习面试题合集Collection of Machine Learning Interview Questions

The Machine Learning part of the interview is usually the most elaborate one. That's the reason we h...

 GarfieldEr007 2016-03-11 14:11:35 阅读数：3735

第一次深度学习实习生面试经历

投了很多简历，只有这一家给了我面试，首先还是比较感谢的。一家创业型公司，不过看到的时候还有有点小吃惊，是一个住宅大厦里面，里面本身有很多公司，进门看上去也是一个住宅的感觉，房子比较小，里面就俩人。。...

 Beach_pants 2017-09-22 09:24:00 阅读数：1466

机器学习——神经网络、深度学习 知识点总结 及 面试题汇总

1、反向传播思想：计算出输出与标签间的损失函数值，然后计算其相对于每个神经元的梯度，根据梯度方向更新权值。（1）将训练集数据输入到ANN的输入层，经过隐藏层，最后达到输出层并输出结果，...

 qq_34896915 2017-06-22 10:55:38 阅读数：8765

为什么网上鲜有炫富的程序员

程序员

百度广告




深度学习：CNN RNN DNN 区别 卷积层或是LSTM单元

梯度消失：<http://www.cnblogs.com/tsiangleo/p/6151560.html> 根本的问题其实并非是消失的梯度问题或者激增的梯度问题，而是在前面的层上的梯度是来自后...

 erinapple 2017-11-04 20:30:04 阅读数：375

文本挖掘--DNN和LSTM的比较

<http://www.cnblogs.com/doublemystery/p/5092014.html> 转载链接

 wang1127248268 2017-08-12 08:56:51 阅读数：496