

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? - 人工智能 - 知乎

zhihu.com Updated Apr 19th, 2016

人工智能 语音识别 图像识别 神经网络 深度学习 (Deep Learning)

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? 以及他们的主要用途是什么? 只知道CNN是局部感受和参数共享, 比较适合用于图像这方面。刚入门的小白真心求助

1 条评论 分
享

12 个回答

按投票排

科研君, 每周六提供一份高质量回答。

卢不群、董豪宇、知乎用户 等人赞同



首先, 我感觉不必像 @李Shawn 同学一样认为DNN、CNN、RNN完全不能相提并论。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, 包括卷积层或是LSTM单元。但是就题主的意思来看, 这里的DNN应该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。因此, 题主一定要将DNN、CNN、RNN等进行对比, 也未尝不可。

其实, 如果我们顺着神经网络技术发展的脉络, 就很容易弄清这几种网络结构发明的初衷, 和他们之间本质的区别, 希望对题主有所帮助。

===== 分割线就是 我
=====

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Rosenblatt。 (扯一个不相关的: 由于计算技术的落后, 当时感知器传输函数是用线拉动变阻器改变电阻的方法机械实现的, 脑补一下科学家们扯着密密麻麻的导线的样子...) 但是, Rosenblatt的单层感知机有一个严重得不能再严重的问题, 即它对稍复杂一些的函数都无能为力 (比如最为典型的“异或”操作)。连异或都不能拟合, 你还能指望这货有什么实际用途么o(′ □ ′)o

随着数学的发展, 这个缺点直到上世纪八十年代才被Rumelhart、Williams、Hinton、LeCun等人 (反正就是一票大牛) 发明的多层感知机 (multilayer perceptron) 克服。多层感知机, 顾名

思义，就是有多层隐含层的感知机（废话……）。好好，我们看一下多层感知机的结构：

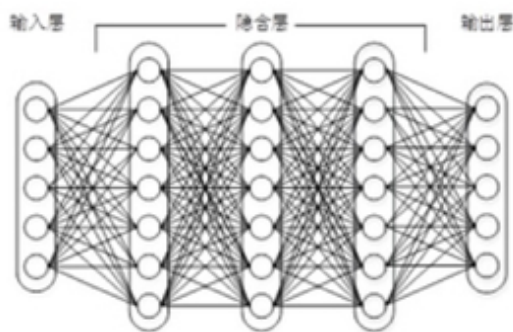


图1上下层神经元全部相连的神经网络——多层感知机

多层感知机可以摆脱早期离散传输函数的束缚，使用sigmoid或tanh等连续函数模拟神经元对激励的响应，在训练算法上则使用Werbos发明的反向传播BP算法。对，这货就是我们现在所说的神经网络NN——神经网络听起来不知道比感知机高端到哪里去了！这再次告诉我们起一个好听的名字对于研（zhuang）究（bi）很重要！

多层感知机解决了之前无法模拟异或逻辑的缺陷，同时更多的层数也让网络更能够刻画现实世界中的复杂情形。相信年轻如Hinton当时一定是春风得意。

多层感知机给我们带来的启示是，神经网络的层数直接决定了它对现实的刻画能力——利用每层更少的神经元拟合更加复杂的函数[1]。

（Bengio如是说：functions that can be compactly represented by a depth k architecture might require an exponential number of computational elements to be represented by a depth $k - 1$ architecture.）

即便大牛们早就预料到神经网络需要变得更深，但是有一个梦魇总是萦绕左右。随着神经网络层数的加深，优化函数越来越容易陷入局部最优解，并且这个“陷阱”越来越偏离真正的全局最优。利用有限数据训练的深层网络，性能还不如较浅层网络。同时，另一个不可忽略的问题是随着网络层数增加，“梯度消失”现象更加严重。具体来说，我们常常使用sigmoid作为神经元的输入输出函数。对于幅度为1的信号，在BP反向传播梯度时，每传递一层，梯度衰减为原来的0.25。层数一多，梯度指数衰减后低层基本上接受不到有效的训练信号。

2006年，Hinton利用预训练方法缓解了局部最优解问题，将隐含层推动到了7层[2]，神经网络真正意义上有了“深度”，由此揭开了深度学习的热潮。这里的“深度”并没有固定的定义——在语音识别中4层网络就能够被认为是“较深的”，而在图像识别中20层以上的网络屡见不鲜。为了克服梯度消失，ReLU、maxout等传输函数代替了sigmoid，形成了如今DNN的基本形式。单从结构上来说，全连接的DNN和图1的多层感知机是没有任何区别的。

值得一提的是，今年出现的高速公路网络（highway network）和深度残差学习（deep residual learning）进一步避免了梯度消失，网络层数达到了前所未有的一百多层（深度残差学习：152层）[3,4]！具体结构题主可自行搜索了解。如果你之前在怀疑是不是有很多方法打上了“深度学习”的噱头，这个结果真是深得让人心服口服。

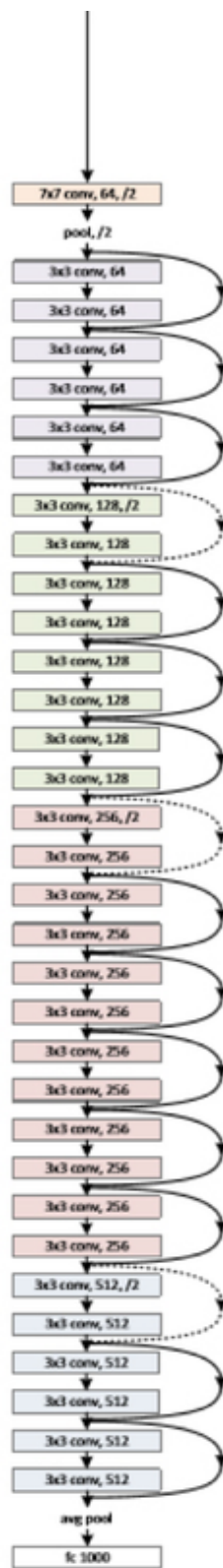


图2缩减版的深度残差学习网络，仅有34层，终极版有152层，自行感受一下

如图1所示，我们看到全连接DNN的结构里下层神经元和所有上层神经元都能够形成连接，带来的潜在问题是参数数量的膨胀。假设输入的是一幅像素为 $1K \times 1K$ 的图像，隐含层有 $1M$ 个节点，光这一层就有 10^{12} 个权重需要训练，这不仅容易过拟合，而且极易陷入局部最优。另外，图像中有固有的局部模式（比如轮廓、边界，人的眼睛、鼻子、嘴等）可以利用，显然应该将图像处理中的概念和神经网络技术相结合。此时我们可以祭出题主所说的卷积神经网络CNN。对于CNN来说，并不是所有上下层神经元都能直接相连，而是通过“卷积核”作为中介。同一个卷积核在所有图像内是共享的，图像通过卷积操作后仍然保留原先的位置关系。两层之

间的卷积传输的示意图如下：

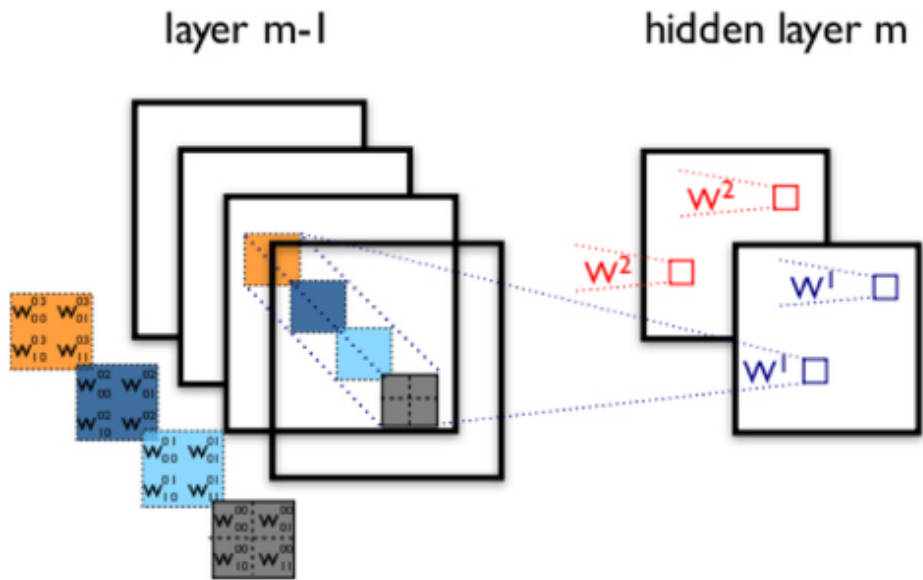


图3卷积神经网络隐含层（摘自Theano教程）

通过一个例子简单说明卷积神经网络的结构。假设图3中 $m-1=1$ 是输入层，我们需要识别一幅彩色图像，这幅图像具有四个通道ARGB（透明度和红绿蓝，对应了四幅相同大小的图像），假设卷积核大小为 100×100 ，共使用100个卷积核 w_1 到 w_{100} （从直觉来看，每个卷积核应该学习到不同的结构特征）。用 w_1 在ARGB图像上进行卷积操作，可以得到隐含层的第一幅图像；这幅隐含层图像左上角第一个像素是四幅输入图像左上角 100×100 区域内像素的加权求和，以此类推。同理，算上其他卷积核，隐含层对应100幅“图像”。每幅图像对是对原始图像中不同特征的响应。按照这样的结构继续传递下去。CNN中还有max-pooling等操作进一步提高鲁棒性。

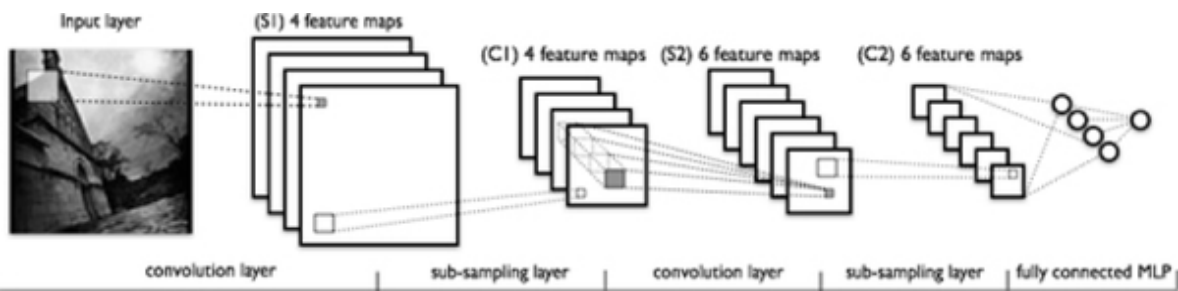


图4一个典型的卷积神经网络结构，注意到最后一层实际上是一个全连接层（摘自Theano教程）

在这个例子里，我们注意到输入层到隐含层的参数瞬间降低到了 $100 \times 100 \times 100 = 10^6$ 个！这使得我们能够用已有的训练数据得到良好的模型。题主所说的适用于图像识别，正是由于CNN模型限制参数了个数并挖掘了局部结构的这个特点。顺着同样的思路，利用语音语谱结构中的局部信息，CNN照样能应用在语音识别中。

全连接的DNN还存在着另一个问题——无法对时间序列上的变化进行建模。然而，样本出现的时间顺序对于自然语言处理、语音识别、手写体识别等应用非常重要。为了适应这种需求，就出现了题主所说的另一种神经网络结构——循环神经网络RNN。

在普通的全连接网络或CNN中，每层神经元的信号只能向上一层传播，样本的处理在各个时刻

独立，因此又被成为前向神经网络(Feed-forward Neural Networks)。而在RNN中，神经元的输出可以在下一个时间戳直接作用到自身，即第i层神经元在m时刻的输入，除了 (i-1) 层神经元在该时刻的输出外，还包括其自身在 (m-1) 时刻的输出！表示成图就是这样的：

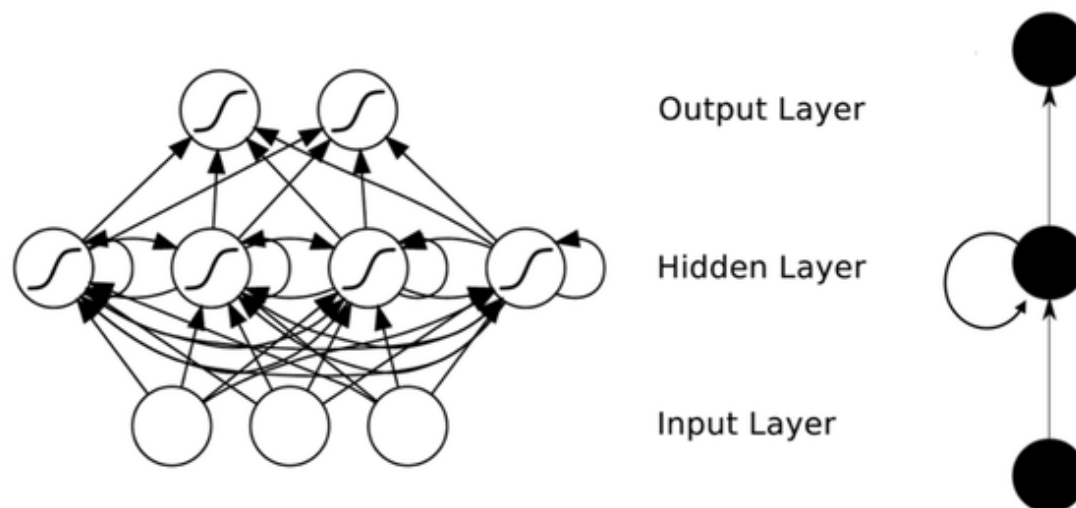


图5 RNN网络结构

我们可以看到在隐含层节点之间增加了互连。为了分析方便，我们常将RNN在时间上进行展开，得到如图6所示的结构：

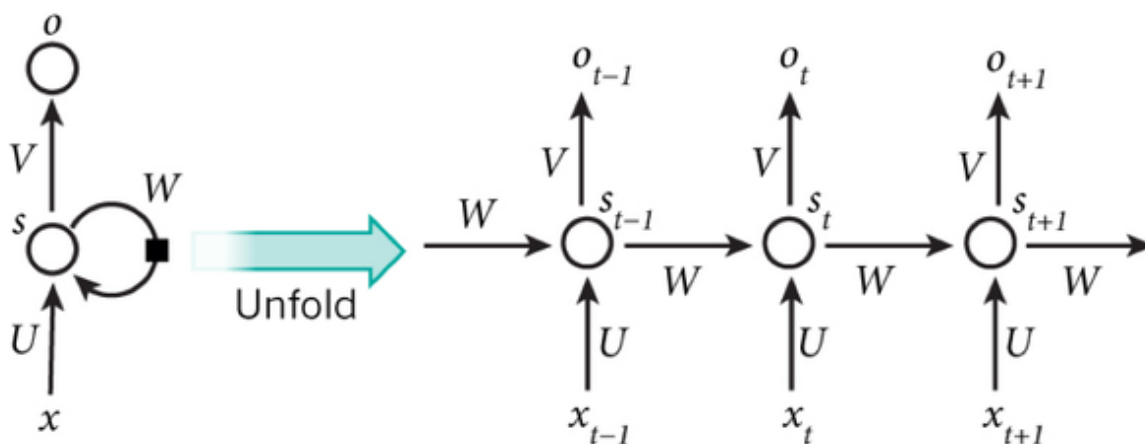


图6 RNN在时间上进行展开

Cool，(t+1) 时刻网络的最终结果O(t+1)是该时刻输入和所有历史共同作用的结果！这就达到了对时间序列建模的目的。

不知题主是否发现，RNN可以看成在一个时间上传递的神经网络，它的深度是时间的长度！正如我们上面所说，“梯度消失”现象又要出现了，只不过这次发生在时间轴上。对于t时刻来说，它产生的梯度在时间轴上向历史传播几层之后就消失了，根本就无法影响太遥远的过去。因此，之前说“所有历史”共同作用只是理想的情况，在实际中，这种影响也就只能维持若干个时间戳。

为了解决时间上的梯度消失，机器学习领域发展出了长短时记忆单元LSTM，通过门的开关实现时间上的梯度传播，防止梯度消失。这个LSTM单元长这个样子：

现时间上记忆功能，开防止梯度消失，一个LSIM单元长这个样子：

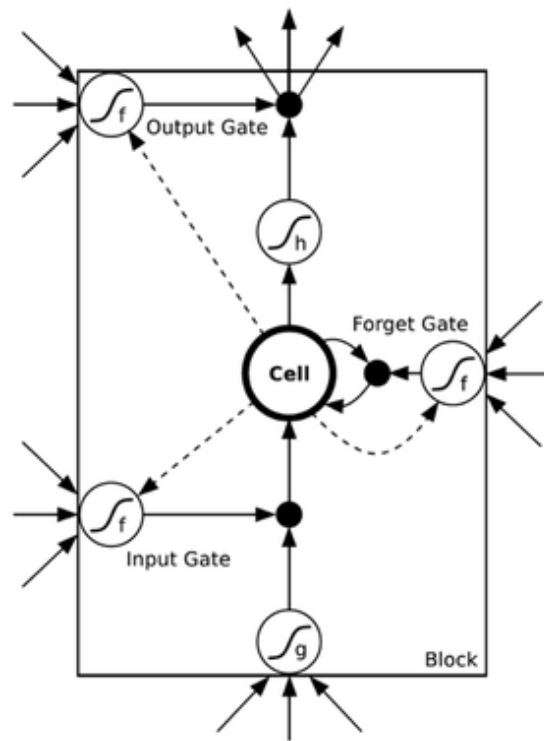


图7 LSTM的模样

除了题主疑惑的三种网络，和我之前提到的深度残差学习、LSTM外，深度学习还有许多其他的结构。举个例子，RNN既然能继承历史信息，是不是也能吸收点未来的信息呢？因为在序列信号分析中，如果我能预知未来，对识别一定也是有所帮助的。因此就有了**双向RNN**、**双向LSTM**，同时利用历史和未来的信息。

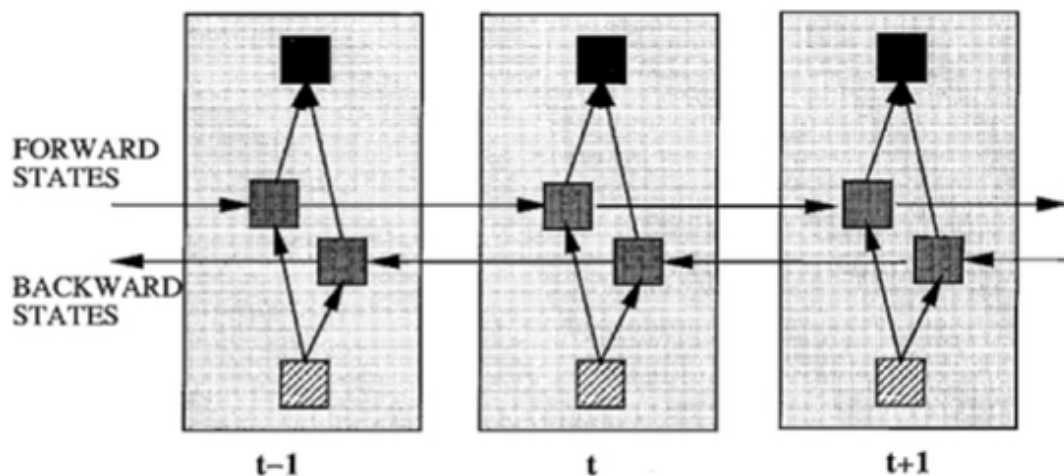


图8双向RNN

事实上，不论是那种网络，他们在实际应用中常常都混合着使用，比如**CNN**和**RNN**在上层输出之前往往会接上全连接层，很难说某个网络到底属于哪个类别。不难想象随着深度学习热度的延续，更灵活的组合方式、更多的网络结构将被发展出来。尽管看起来千变万化，但研究者们出发点肯定都是为了解决特定的问题。题主如果想进行这方面的研究，不妨仔细分析一下这些结构各自的特点以及它们达成目标的手段。入门的话可以参考：

Ng写的Ufidl: [UFLDL教程 - Ufidl](#)

也可以看Theano内自带的教程, 例子非常具体: [Deep Learning Tutorials](#)

欢迎大家继续推荐补充。

当然啦, 如果题主只是想凑个热闹时髦一把, 或者大概了解一下方便以后把妹使, 这样看看也就罢了吧。

参考文献:

[1] Bengio Y. Learning Deep Architectures for AI[J]. Foundations & Trends® in Machine Learning, 2009, 2(1):1-127.

[2] Hinton G E, Salakhutdinov R R. Reducing the Dimensionality of Data with Neural Networks[J]. Science, 2006, 313(5786):504-507.

[3] He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. arXiv:1512.03385, 2015.

[4] Srivastava R K, Greff K, Schmidhuber J. Highway networks. arXiv:1505.00387, 2015.

【非常高兴看到大家喜欢并赞同我们的回答。应许多知友的建议, 最近我们开通了同名公众号: **PhDer**, 也会定期更新我们的文章, 如果您不想错过我们的每篇回答, 欢迎扫码关注~】



编辑于 2016-03-23 40 条评论 感谢 分享 收藏 • 没有帮助 • 举报 • 禁止转载

李Shawn，人工智能研究人员/学生



知乎用户、白海江、ClairBoyanK 等人赞同

个人觉得CNN、RNN和DNN不能放在一起比较。

DNN是一个大类，CNN是一个典型的空间上深度的神经网络，RNN是在时间上深度的神经网络。

推荐你从UFLDL开始看，这是斯坦福深度学习的课程，了解一些神经网络的基础，会对你的学习有很大帮助。

=====分割线

前面一位同学回答得非常详细完整，我再回来谈一谈怎么学习这些模型，我来分享一下我的学习历程。我也是在学习过程中，以后会慢慢继续补充。

1、ufldl.stanford.edu/wiki... 教程

这是我最开始接触神经网络时看的资料，把这个仔细研究完会对神经网络的模型以及如何训练（反向传播算法）有一个基本的认识，算是一个基本功。

2、Deep Learning Tutorials

这是一个开源的深度学习工具包，里面有很多深度学习模型的python代码还有一些对模型以及代码细节的解释。我觉得学习深度学习光了解模型是不难的，难点在于把模型落地写成代码，因为里面会有很多细节只有动手写了代码才会了解。但Theano也有缺点，就是极其难以调试，以至于我后来就算自己动手写几百行的代码也不愿意再用它的工具包。所以我觉得Theano的正确用法还是在于看里面解释的文字，不要害怕英文，这是必经之路。PS：推荐使用python语言，目前来看比较主流。

3、Stanford University CS231n: Convolutional Neural Networks for Visual Recognition

斯坦福的一门课：卷积神经网络，李飞飞教授主讲。这门课会系统的讲一下卷积神经网络的模型，然后还有一些课后习题，题目很有代表性，也是用python写的，是在一份代码中填写一部分缺失的代码。如果把这个完整学完，相信使用卷积神经网络就不是一个大问题了。

4、斯坦福大学公开课：机器学习课程

这可能是机器学习领域最经典最知名的公开课了，由大牛Andrew Ng主讲，这个就不仅仅是深度学习了，它是带你领略机器学习领域中最重要概念，然后建立起一个框架，使你对机器学习这个学科有一个较为完整的认识。这个我觉得所有学习机器学习的人都应该看一下，我甚至在某公司的招聘要求上看到过：认真看过并深入研究过Andrew Ng的机器学习课程，由此可见其重要性。

编辑于 2016-03-07 8 条评论 感谢 分享 收藏 • 没有帮助 • 举报 • 作者保留权利

知乎用户，我是机器鼓励师



李秀丽、金尚赟、与析 赞同

在序列信号的应用上，CNN是只响应预先设定的信号长度（输入向量的长度），RNN的响应长度是学习出来的。

度定'字'习出米的

发布于 2015-08-22 添加评论 感谢 分享 收藏 • 没有帮助 • 举报 • 作者保留权利

知乎用户, Working on RL

堃堃、知乎用户、黄成炎 等人赞同



DNN以神经网络为载体，重在深度，可以说是一个统称。

RNN，回归型网络，用于序列数据，并且有了一定的记忆效应，辅之以Istm。

CNN应该侧重空间映射，图像数据尤为贴合此场景。

发布于 2015-08-31 添加评论 感谢 分享 收藏 • 没有帮助 • 举报 • 作者保留权利

魏秀参，铁打的人儿，流水的饭

知乎用户、张弛、崔强 等人赞同



蟹妖，可参阅我关于深度学习入门的回答。

[深度学习入门必看的书和论文？有哪些必备的技能需学习？ - 魏秀参的回答](#)

编辑于 2015-08-20 添加评论 感谢 分享 收藏 • 没有帮助 • 举报 • 作者保留权利

知乎用户, Shallow Learning

arisecbf、潇洒小妞、俗人 赞同



中文不好解释，用英文试试，不当之处请谅解。

Artificial neural networks use networks of activation units (hidden units) to map inputs to outputs. The concept of deep learning applied to this model allows the model to have multiple layers of hidden units where we feed output from the previous layers. However, dense connections between the layers is not efficient, so people developed models that perform better for specific tasks.

The whole "convolution" in convolutional neural networks is essentially based on the fact that we're lazy and want to exploit spatial relationships in images. This is a huge deal because we can then group small patches of pixels and effectively "downsample" the image while training multiple instances of small detectors with those patches. Then a CNN just moves those filters around the entire image in a convolution. The outputs are then collected in a pooling layer. The pooling layer is again a down-sampling of the previous feature map. If we have activity on an output for filter a, we don't necessarily care if the activity is for (x, y) or (x+1, y), (x, y+1) or (x+1, y+1), so long as we have activity. So we often just take the highest value of activity on a small grid in the feature map called max pooling.

If you think about it from an abstract perspective, the convolution part of a CNN is effectively doing a reasonable way of dimensionality reduction. After a while you can flatten the image and process it through layers in a dense network. Remember to use dropout layers! (because our guys wrote that paper :P)

I won't talk about RNN for now because I don't have enough experience working with them and according to my colleague nobody really knows what's going on inside an LSTM...but they're insanely popular for data with time dependencies.

Let's talk RNN. Recurrent networks are basically neural networks that evolve through time.

Rather than exploiting spatial locality, they exploit sequential, or temporal locality. Each iteration in an RNN takes an input and its previous hidden state, and produces some new hidden state. The weights are shared in each level, but we can unroll an RNN through time and get your everyday neural net. Theoretically RNN has the capacity to store information from as long ago as possible, but historically people always had problems with the gradients vanishing as we go back further in time, meaning that the model can't be differentiated numerically and thus cannot be trained with backprop. This was later solved in the proposal of the LSTM architecture and subsequent work, and now we train RNNs with BPTT (backpropagation through time). Here's a link that explains LSTMs really well: [colah.github.io/posts/2...](https://colah.github.io/posts/2015-08-Understanding-Recurrent-Neural-Networks/)

Since then RNN has been applied in many areas of AI, and many are now designing RNN with the ability to extract specific information (read: features) from its training examples with attention-based models.

编辑于 2016-02-15 1 条评论 感谢 分享 收藏 • 没有帮助 • 举报 • 作者保留权利

知乎用户，梅吹，老司机。

白海江、Dirge、骑鱼的基德 等人赞同

CNN就是全连接权值太多，取的一个折衷策略，只取部分连接边扫描一遍整个输入，最后再汇总（求max等等操作）。

RNN就是在隐层加入了自连边和互连边（即隐层可以相互连接），可以按时序展开为一系列FNN。常见训练算法为bptt和lstm。

DNN个人理解就是隐层有很多层，bp反传时可能梯度会锐减或剧增，导致误差传不回来，可以通过重新设计网络结构（类似lstm）的办法来解决。

发布于 2015-08-18 3 条评论 感谢 分享 收藏 • 没有帮助 • 举报 • 禁止转载

知乎用户，计算机专业

丁维悦、李卓 赞同

我个人的看法是，DNN是深度学习思想的一个统称，狭义上它是一种逐层贪婪无监督的学习方法，注意它只是一种思想，需要依赖具体的模型去实现，即深度学习 = 逐层贪婪无监督方法 + 特定深度学习模型。而具体的深度学习模型有CNN、RNN、SDA、RBM、RBM等。逐层贪婪无监督的深度学习思想是2006年提出来的，能有效解决深层网络结构训练的梯度弥散等问题。而CNN等网络模型其实很早就提出来了，之前用BP之类的方法效果一直不太好，等到深度学习思想提出后，这些老模型就都焕发了新生，带来了今天深度学习的火热。回归到问题上来，就是CNN、RNN、DNN不是一个层次上的东西，分别是具体的模型和通用的思想。这是我的个人理解，还在学习中，如有错误欢迎指出。

编辑于 2016-02-13 1 条评论 感谢 分享 收藏 • 没有帮助 • 举报 • 作者保留权利

知乎用户，大王派我来巡山。

梁正中、puyangsky、知乎用户 赞同

想学专业知识就要少上知乎 多看书

编辑于 2016-02-15 添加评论 感谢 分享 收藏 • 没有帮助 • 举报 • 作者保留权利

Gundam, The truth is what it is, not what you c

DNN区别于浅层神经网络，是所有深度学习中的网络模型统称，包括CNN、RNN等，以后还会有新的网络模型提出。题主说的CNN应该是三大特点：local receptive fields、shared weights和pooling。他们的用途都是拿数据来训练学习特征（太笼统了，嘿嘿），省去手工提取特征的过程，类似是一个通用的学习框架。[@helloworld00](#)回答中提到的课程不错，针对卷积神经网络和视觉识别的，题主可以深入学习一下。UFLDL也很适合初学者。另外如果真想研究深度学习，还是要看论文的-->[Reading List « Deep Learning](#)

编辑于 2016-02-28 添加评论 感谢 分享 收藏 • 没有帮助 • 举报 • 作者保留权利

俞海宝, 计算机视觉爱好者/文艺小青年

祝三 赞同

请问现在将CNN中的feature map可视化有哪些思路和方法吗？

发布于 2016-03-18 添加评论 感谢 分享 收藏 • 没有帮助 • 举报 • 作者保留权利

知乎用户, 朝闻道，夕也要闻道。

李尧 赞同

理解概念简单，代码实现落地难啊。

发布于 2016-03-24 添加评论 感谢 分享 收藏 • 没有帮助 • 举报 • 作者保留权利

我来回答这个问题

写回答...

我要回答

