Convolutional Neural Networks卷积神经网络

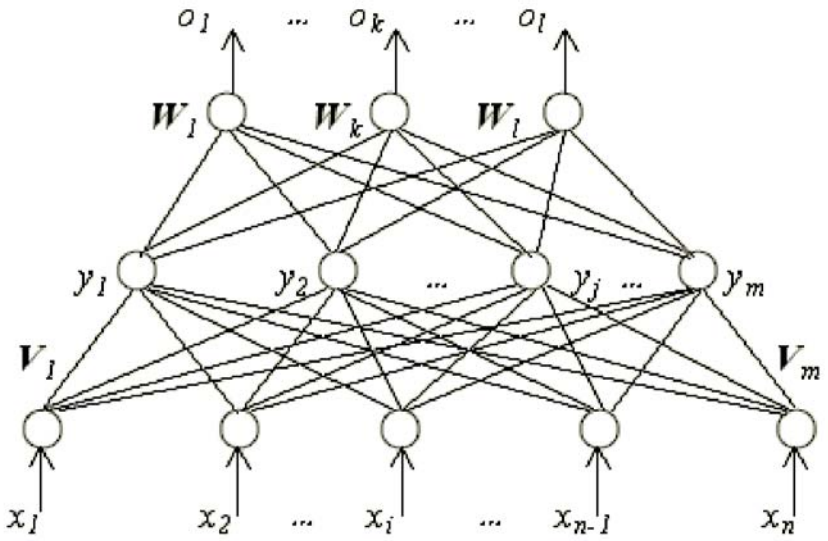
Contents

本文是我在20140822的周报，其中部分参照了以下博文或论文，如果在文中有一些没说明白的地方，可以查阅他们。对Yann LeCun前辈，和celerychen2009、zouxy09表示感谢。

1. Deep Learning（深度学习）学习笔记整理系列之（七）
2. Deep Learning论文笔记之（四）CNN卷积神经网络推导和实现
3. 卷积神经网络
4. 反向传导算法
5. Yann LeCun's Publications 中1998年著名的"Gradient-Based Learning Applied to Document Recognition"
6. 反向传播BP算法

一：前导 **Back Propagation**反向传播算法

网络结构



经典的BP网络是三层结构：输入层X、输出层O和隐层Y。

输入向量：X = (x1,x2,…,xn)T

隐层输出：Y = (y1,y2,…,ym)T　权值
V = (v1,v2,…,vm)T

输出向量：O = (o1,o2,…,ol)T　权值W = (w1,w2,…,wl)T

期望输出：D = (d1,d2,…,dn)T

学习算法

输入层到隐层的计算过程：

$$net_j = \sum_{i=0}^{i=n} v_{ji} x_j$$

$$y_j = f(net_j), \quad 其中, \quad j = 1,2,...,m$$

隐层到输出层的计算过程：

$$net_k = \sum_{j=0}^{j=m} w_{kj} y_j$$

$$o_k = f(net_k)，\text{其中} k = 1,2,...,l$$

网络输出层误差函数为：

$$E = \frac{1}{2}(d-o)^2 = \frac{1}{2}\sum_{k=1}^{l}(d_k - o_k)^2$$

将误差函数展开到隐层，则为：

$$E = \frac{1}{2}(d-o)^2 = \frac{1}{2}\sum_{k=1}^{l}(d_k - o_k)^2 = \frac{1}{2}\sum_{k=1}^{l}(d_k - f(\sum_{j=0}^{j=m} w_{kj} y_j))^2 = \frac{1}{2}\sum_{k=1}^{l}(d_k - f(\sum_{j=0}^{j=m} w_{kj} f(\sum_{i=0}^{i=n} v_{ji} x_j)))^2$$

训练过程就是要让最后的E减到尽可能小，以达到最优值，所以可以将E对每一个输入参数求偏导，以达到最优。所以：

$$\Delta w_{kj} = -\eta \frac{\partial E}{\partial w_{kj}}, k = 1,2,...,l; j = 0,1,2,...,m$$

$$\Delta v_{ji} = -\eta \frac{\partial E}{\partial v_{ji}}, i = 0,1,2,...,n; j = 1,2,...,m$$

η是一个比例系数，经过一系列计算之后，上式可化成：

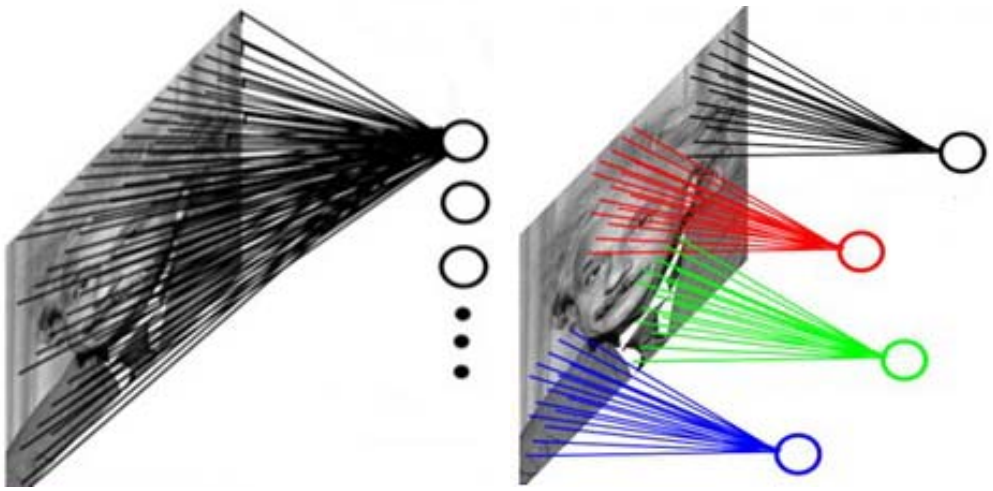
$$\Delta w_{kj} = \eta \delta_k^o y_j = \eta (d_k - o_k) o_k (1 - o_k) y_j$$

$$\Delta v_{ji} = \eta \delta_j^y x_i = \eta (\sum_{k=1}^{l} \delta_k^o w_{jk}) y_j (1 - y_j) x_i$$

通过极小化误差反向传播调整权值矩阵，不断循环直到最佳。

二：**Convolutional Neural Networks**卷积神经网络

BP神经网络每一层节点是一个线性的一维排列状态，层与层的网络节点之间是全连接的。而如果BP网络中层与层之间的节点连接不再是全连接，而是局部连接的。这样，就是一种最简单的一维卷积网络。如果我们把上述这个思路扩展到二维，这就是我们在大多数参考资料上看到的卷积神经网络，具体参看图2：



a.全连接网络　　b.局部连接网络

图2

图2.a：全连接网络。如果L1层有1000×1000像素的图像，L2层有1000,000个隐层神经元，每个隐层神经元都连接L1层图像的每一个像素点，就有1000x1000x1000,000=10^12个连接，也就是10^12个权值参数。

图2.b：局部连接网络。L2层每一个节点与L1层节点同位置附近10×10的窗口相连接，则1百万个隐层神经元就只有100w乘以100，即10^8个参数。其权值连接个数比原来减少了四个数量级。

卷积神经网络另外一个特性是权值共享。例如，就图2.b来说，权值共享，不是说，所有的红色线标注的连接权值相同。而是说，每一个颜色的线都有一个红色线的权值与之相等，所以第二层的每个节点，其从上一层进行卷积的参数都是相同的。

图2中隐层的每一个神经元都连接10×10个图像区域，也就是说每一个神经元存在10×10=100个连接权值参数。如果我们每个神经元这100个参数是相同的？也就是说每个神经元用的是同一个卷积核去卷积图像。这样L1层我们就只有100个参数。但是这样，只提取了图像一种特征？如果需要提取不同的特征，就加多几种卷积核。所以假设我们加到100种卷积核，也就是1万个参数。

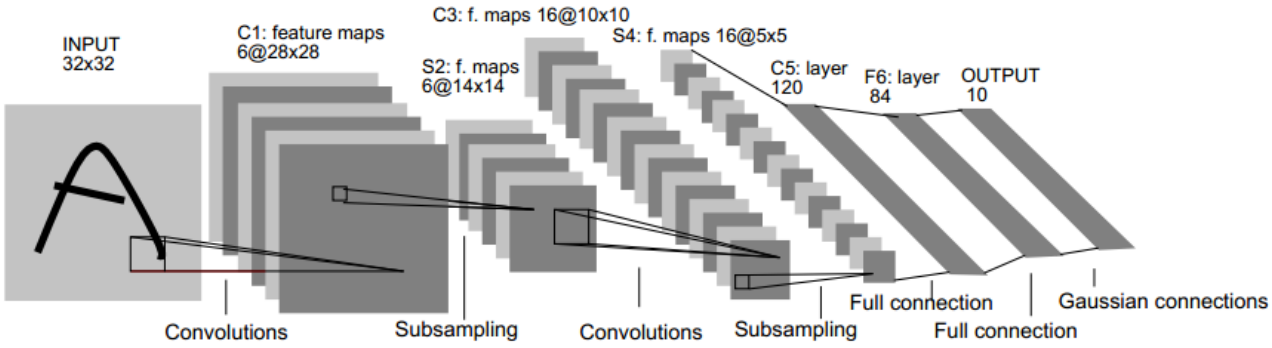
每种卷积核的参数不一样，表示它提出输入图像的不同特征（不同的边缘）。这样每种卷积核去卷积图像就得到对图像的不同特征的放映，我们称之为Feature Map，也就是特征图。

需要注意的一点是，上面的讨论都没有考虑每个神经元的偏置，加上偏置参数，则每个神经元需要的权值参数个数需要加1。

上面描述的只是单层网络结构，Yann LeCun等在1998年发布的论文“Gradient-Based Learning Applied to Document Recognition”提出了基于卷积神经网络的一个文字识别系统 LeNet-

5，随后被用于银行手写数字的识别。

三：LeCun的LeNet-5



不包含输入，LeNet-5共有7层，每层都包含连接权值（可训练参数）。输入图像为32*32大小。我们先要明确一点：每个层有多个特征图，每个特征图通过一种卷积滤波器提取输入的一种特征，然后每个特征图有多个神经元。

C1、C3、C5是卷积层，S2、S4、S6是下采样层。利用图像局部相关性的原理，对图像进行下抽样，可以减少数据处理量同时保留有用信息。

C1层是卷积层，由6个特征图构成。特征图中每个神经元与输入层中5*5的邻域相连。C1的大小为28*28，这样能防止输入的连接掉到边界之外。C1有156个可训练参数，共122,304个连接。

可训练参数是卷积核可训练参数个数加上一个偏置参数，再乘以特征图个数。（公式）

连接数即可训练参数乘以特征图大小。（公式）

对于C1：

(5*5+1)*6=156个参数

156*(28*28)=122,304个连接

S2层是下采样层，有6个14*14的特征图。特征图中的每个单元与C1中相对应特征图的2*2邻域相连接。S2层每个单元的4个输入相加，乘以一个可训练参数，再加上一个可训练偏置。结果通过sigmoid函数计算。可训练系数和偏置控制着sigmoid函数的非线性程度。如果系数比较小，那么运算近似于线性运算，亚采样相当于模糊图像。如果系数比较大，根据偏置的大小亚采样可以被看成是有噪声的“或”运算或者有噪声的“与”运算。每个单元的2*2感受野并不重叠，

对于S2：

每个特征图的大小是C1中特征图大小的1/4（行和列各1/2）。S2层有(1+1)*6=12个可训练参数和14*14*(4+1)*6=5880个连接。

C3层也是一个卷积层，它同样通过5×5的卷积核去卷积层S2，然后得到的特征map就只有10×10个神经元，但是它有16种不同的卷积核，所以就存在16个特征图了。C3中的每个特征map是连接到S2中的所有6个或者几个特征map的，表示本层的特征map是上一层提取到的特征map的不同组合（这个做法也并不是唯一的）。

为什么不把S2中的每个特征图连接到每个C3的特征图呢？原因有2点。第一，不完全的连接机制将连接的数量保持在合理的范围内。第二，其破坏了网络的对称性。由于不同的特征图有不同的输入，所以迫使他们抽取不同的特征。

LeCun采用的方式是：C3的前6个特征图以S2中3个相邻的特征图子集为输入。接下来6个特征图以S2中4个相邻特征图子集为输入。然后的3个以不相邻的4个特征图子集为输入。最后一个将S2中所有特征图为输入。如图：

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	X				X	X	X			X	X	X	X		X	X
1	X	X				X	X	X			X	X	X	X		X
2	X	X	X				X	X	X			X		X	X	X
3		X	X	X			X	X	X	X			X		X	X
4			X	X	X			X	X	X	X		X	X		X
5				X	X	X			X	X	X	X		X	X	X

这样C3层有(25*3+1)*6+(25*4+1)*6+(25*4+1)*3+(25*6+1)*1=1516个可训练参数和((25*3+1)*6+(25*4+1)*6+(25*4+1)*3+(25*6+1)*1)*(10*10)=151600个连接。

S4层是下采样层，由16个5*5大小的特征图构成。特征图中的每个单元与C3中相应特征图的2*2邻域相连接，跟C1和S2之间的连接一样。S4层有16*(1+1)=32个可训练参数（每个特征图1个因子和一个偏置）和5*5*(4+1)*16=2000个连接。（如果到这看不懂公式，可以不按照顺序阅读，先看所有的卷积层，再看所有的下采样层）

C5层是一个卷积层，有120个特征图。每个单元与S4层的全部16个单元的5*5邻域相连。由于S4层特征图的大小也为5*5（同滤波器一样），故C5特征图的大小为1*1：这构成了S4和C5之间的全连接。

之所以仍将**C5**标示为卷积层而非全相联层，是因为如果**LeNet-5**的输入变大，而其他的保持不变，那么此时特征图的维数就会比**1*1**大。C5层有(5*5*16+1)*120=48120个可训练参数，由于C5特征图大小为1:1，所以有48120*1*1=48120个链接（Yann原文只是说有48120个trainable connection，与上文所用术语不一样，暂且认为就是说48120个trainable parameters and 48120个connection吧，这个与我们计算出来也一

致）。

F6层有84个单元（之所以选这个数字的原因来自于输出层的设计），与C5层全相连。如同经典神经网络，F6层计算输入向量和权重向量之间的点积，再加上一个偏置。然后将其传递给sigmoid函数产生单元i的一个状态。有(120+1)*84=10164个可训练参数，也是10164个连接。

最后，输出层由欧式径向基函数（*Euclidean Radial Basis Function*）单元组成，每类一个单元，每个有*84*个输入。换句话说，每个输出*RBF*单元计算输入向量和参数向量之间的欧式距离。输入离参数向量越远，*RBF*输出的越大。一个*RBF*输出可以被理解为衡量输入模式和与*RBF*相关联类的一个模型的匹配程度的惩罚项。用概率术语来说，*RBF*输出可以被理解为*F6*层配置空间的高斯分布的负*log-likelihood*。给定一个输入模式，损失函数应能使得*F6*的配置与*RBF*参数向量（即模式的期望分类）足够接近。这些单元的参数是人工选取并保持固定的（至少初始时候如此）。这些参数向量的成分被设为*-1*或*1*。虽然这些参数可以以*-1*和*1*等概率的方式任选，或者构成一个纠错码，但是被设计成一个相应字符类的*7*12*大小（即*84*）的格式化图片。这种表示对识别单独的数字不是很有用，但是对识别可打印*ASCII*集中的字符串很有用。

使用这种分布编码而非更常用的*“1 of N”*编码用于产生输出的另一个原因是，当类别比较大的时候，非分布编码的效果比较差。原因是大多数时间非分布编码的输出必须为*0*。这使得用*sigmoid*单元很难实现。另一个原因是分类器不仅用于识别字母，也用于拒绝非字母。使用分布编码的*RBF*更适合该目标。因为与*sigmoid*不同，他们在输入空间的较好限制的区域内兴奋，而非典型模式更容易落到外边。

*RBF*参数向量起着*F6*层目标向量的角色。需要指出这些向量的成分是*+1*或*-1*，这正好在*F6 sigmoid*的范围内，因此可以防止*sigmoid*函数饱和。实际上，*+1*和*-1*是*sigmoid*函数的最大弯曲的点处。这使得*F6*单元运行在最大非线性范围内。必须避免*sigmoid*函数的饱和，因为这将会导致损失函数较慢的收敛和病态问题。

四：**CNNs**的训练过程

CNNs训练算法与传统的BP算法差不多。主要包括4步，这4步被分为两个阶段：

第一阶段，向前传播阶段：

a）从样本集中取一个样本(X,Yp)，将X输入网络；

b）计算相应的实际输出Op。

在此阶段，信息从输入层经过逐级的变换，传送到输出层。这个过程也是网络在完成训练后正常运行时执行的过程。在此过程中，网络执行的是计算（实际上就是输入与每层的权值矩阵相点乘，得到最后的输出

结果）：

Op=Fn（…（F2（F1（XpW（1））W（2））…）W（n））

第二阶段，向后传播阶段

a）算实际输出Op与相应的理想输出Yp的差；

b）按极小化误差的方法反向传播调整权矩阵。

五：总结

CNNs这种算法目前在图像识别和处理应用的相当广泛。在ImageNet 2014大规模视觉识别比赛中，CNNs得到了广泛的应用，其中错误率只有6.656%的最优算法也是源自CNNs。

Yann LeCun在90年代做出LeNet，到今天成为视觉识别的最主要技术，一方面与他的努力分不开，另一反面，他愿意在神经网络低谷时期坚持研究这个方向，本身也是一直值得学习的精神。

PV：9,669 时间：2014-08-25 [http://www.gageet.com/2014/0878.php] 分类：Deep Learning TAG：BP算法、CNNs、LeNet、weekly report、Yann Lecun、卷积神经网络

网站未在畅言补全备案信息，当前无法使用评论服务，请联系网站管理员。

评论(0人参与，0条评论)

不需登陆ye可评论

微博登录 QQ登录 游客

还没有评论，快来抢沙发吧！