

资讯专栏

INFORMATION COLUMN

春季采购钜惠来袭 | 全球大促 · 爆款云产品 · 低至0.9折!

免费试用

首页 / 文章专栏 / 深度学习教父Geoffrey Hinton的“胶囊理论”终于发出论文

深度学习教父Geoffrey Hinton的“胶囊理论”终于发出论文

VincentFF 发布于2019-04-25 18:20 / 3419人阅读

摘要：在底层的胶囊之后连接了层和层。胶囊效果的讨论在论文最后，作者们对胶囊的表
们认为，由于胶囊具有分别处理不同属性的能力，相比于可以提高对图像变换的健壮性，
有出色的表现。



背景

目前的神经网络中，每一层的神经元都做的是类似的事情，比如一个卷积层内的每个神经元都做的是一样的
卷积操作。而Hinton坚信，不同的神经元完全可以关注不同的实体或者属性，比如在一开始就有不同的神经
元关注不同的类别（而不是到最后才有归一化分类）。具体来说，有的神经元关注位置、有的关注尺寸、有
的关注方向。这类似人类大脑中语言、视觉都有分别的区域负责，而不是分散在整个大脑中。

为了避免网络结构的杂乱无章，Hinton提出把关注同一个类别或者同一个属性的神经元打包集合在一起，好
像胶囊一样。在神经网络工作时，这些胶囊间的通路形成稀疏激活的树状结构（整个树中只有部分路径上的
胶囊被激活），从而形成了他的Capsule理论。值得一提的是，同在谷歌大脑（但不在同一个办公室）的Jeff
Dean也认为稀疏激活的神经网络是未来的重要发展方向，不知道他能不能也提出一些不同的实现方法来。

Capsule这样的网络结构在符合人们“一次认知多个属性”的直观感受的同时，也会带来另一个直观的问题，
那就是不同的胶囊应该如何训练、又如何让网络自己决定胶囊间的激活关系。Hinton这篇论文解决的重点问
题就是不同胶囊间连接权重（路由）的学习。

点击立即咨询

目元，每个层中的神经元组合形成不同的胶囊，每个胶囊有一个活动向量（activity vector），它是这个胶囊对于它关注的类别或者属性的表征。树结构中的每个节点就对应着一个活动的胶囊。通过一个迭代路由的过程，每个活动的胶囊都会从高一层网络中的胶囊中选择一个，让它成为自己的母节点。对于高阶的视觉系统来说，这样的迭代过程就很有潜力解决一个物体的部分如何层层组合成整体的问题。

对于实体在网络中的表征，众多属性中有一个属性比较特殊，那就是它出现的概率（网络检测到某一类物体出现的置信度）。一般典型的方式是用一个多带带的、输出0到1之间的回归单元来表示，0就是没出现，1就是出现了。在这篇论文中，Hinton想用活动向量同时表示一个实体是否出现以及这个实体的属性。他的做法是用向量不同维度上的值分别表示不同的属性，然后用整个向量的模表示这个实体出现的概率。为了保证向量的长度，也就是实体出现的概率不超过1，向量会通过一个非线性计算进行标准化，这样实体的不同属性也就实际上体现为了这个向量在高维空间中的方向。

采用这样的活动向量有一个很大的好处，就是可以帮助低层级的胶囊选择自己连接到哪个高层级的胶囊。具体做法是，一开始低层级的胶囊会给所有高层级的胶囊提供输入；然后这个低层级的胶囊会把自己的输出和一个权重矩阵相乘，得到一个预测向量。如果预测向量和某个高层级胶囊的输出向量的标量积更大，就可以形成从上而下的反馈，提高这两个胶囊间的耦合系数，降低低层级胶囊和其它高层级胶囊行几次迭代后，贡献更大的低层级胶囊和接收它的贡献的高层级胶囊之间的连接就会占越

在论文作者们看来，这种“一致性路由”（routing-by-agreement）的方法要比之前较一个最活跃的特征的路由方法有效得多。

网络构建

作者们构建了一个简单的CapsNet。除最后一层外，网络的各层都是卷积层，但它们现在层，其中用向量输出代替了CNN的标量特征输出、用一致性路由代替了较大池化。与CNN网络观察了图像中更大的范围，不过由于不再是较大池化，所以位置信息一直都得到了保留空间位置的判断也只需要看是哪些胶囊被激活了。



这个网络中最底层的多维度胶囊结构就展现出了不同的特性，它们起到的作用就像传统计算机图形渲染中的不同元素一样，每一个胶囊关注自己的一部分特征。这和目前的计算机视觉任务中，把图像中不同空间位置的元素组合起来形成整体理解（或者说图像中的每个区域都会首先激活整个网络然后再进行组合）具有截然不同的计算特性。在底层的胶囊之后连接了PrimaryCaps层和DigitCaps层。

实验结果

由于胶囊具有新的特性，所以文中的实验结果也并不只是跑跑Benchmark而已，还有很多对胶囊带来的新特性的分析。

数字识别

首先在MNIST数据集上，经过三次路由迭代学习、层数也不算多的CapsNet就得到了优秀的错误率。

(竖线左侧) , CapsNet可以正确识别到图像中的细节, 同时降低噪声。

[页](#) [动](#) [品](#) [案](#) [栏](#) [答](#) [院](#) [家](#)



健壮性

由于网络结构中DigitCaps部分能够分别学到书写中旋转、粗细、风格等变化, 所以对小变化的健壮性更好。在用一个随机抹黑过数字的MNIST数据集训练CapsNet后, 作者们用它来识别affNIST数据集。这个数据集中的样本都是经过小幅度变化后的MNIST样本, 变化后的样本如下图。这个CapsNet直接拿来识别affNIST的正确率有79%; 同步训练的、参数数目类似的CNN只有66%。



分割高度重合的数字

作者们把MNIST数据集中的数字两两叠在一起建立了MultiMNIST数据集, 两个数字的边是重合的。CapsNet的识别结果高于CNN基准自不必提, 但作者们接下来做的图形分析中囊的妙处。

如图, 作者们把两个激活程度较高的胶囊对应的数字作为识别结果, 据此对识别到的图像对于下图中识别正确的样本 (L指真实标签, R指激活程度较高的两个胶囊对应的标签), 的胶囊各自工作, 在一个识别结果中用到的特征并不会影响到另一个识别结果, 不受重叠叠部分的特征可以复用) 。



另一方面, 每个胶囊还是需要足够多的周边信息支持, 而不是一味地认为重叠部分的特征左图是选了一个高激活程度的胶囊和一个低激活程度胶囊的结果 (*R表示其中一个数字既是识别结果, L仍然为真实标签)。可以看到, 在 (5, 0) 图中, 关注 “7” 的胶囊并没有找到足够多的 “7” 的特征, 所以激活很弱; (1, 8) 图中也是因为没有 “0” 的支持特征, 所以重叠的部分也没有在 “0” 的胶囊中用第二次。



胶囊效果的讨论

在论文最后, 作者们对胶囊的表现进行了讨论。他们认为, 由于胶囊具有分别处理不同属性的能力, 相比于CNN可以提高对图像变换的健壮性, 在图像分割中也会有出色的表现。胶囊基于的 “图像中同一位置至多只有某个类别的一个实体” 的假设也使得胶囊得以使用活动向量这样的分离式表征方式来记录某个类别实例的各方面属性, 还可以通过矩阵乘法建模的方式更好地利用空间信息。不过胶囊的研究也才刚刚开始, 他们觉得现在的胶囊至于图像识别, 就像二十一世纪初的RNN之于语音识别 —— 研究现在只是刚刚起步, 日后定会大放异彩。

论文全文参见: <https://arxiv.org/pdf/1710.09829.pdf>

[点击立即咨询](#)